UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ASSIMILATION OF RADAR AND SURFACE OBSERVATIONS OF A DEVELOPING CONVECTIVE SYSTEM: OBSERVING SYSTEM SIMULATION AND REAL-DATA EXPERIMENTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

RYAN ALAN SOBASH
Norman, Oklahoma
2013

ASSIMILATION OF RADAR AND SURFACE OBSERVATIONS OF A DEVELOPING CONVECTIVE SYSTEM: OBSERVING SYSTEM SIMULATION AND REAL-DATA EXPERIMENTS

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

By

Dr. David Stensrud

Prof. Ming Xue

Prof. Frederick Carr

Prof. Xuguang Wang

Dr. Louis Wicker

Prof. S. Lakshmivarahan

Acknowledgements

The work presented in this dissertation has been improved through collaborations and discussions with a variety of individuals within the University of Oklahoma School of Meteorology, the National Severe Storms Laboratory (NSSL), and the National Center for Atmospheric Research (NCAR). I am particularly grateful for my adviser, Dr. David Stensrud, for supporting and guiding me during the past 4 years. Whenever I became consumed with the details of a particular idea or issue, he always provided a valuable perspective on the larger impact of my work. I've been fortunate to have an adviser and Ph. D. committee that have been willing to give me the flexibility to independently develop and enact a plan of research during the course of my degree. This was, at times, extremely intimidating, but their guidance ensured I stayed on the proper trajectory. This allowed me to experience first-hand the joys, along with the inevitable disappointments, associated with scientific research and the collaborative environment in which it takes place.

Conversations with many scientists at NSSL, including Dr. Louis Wicker, Dr. Michael Coniglio, Dr. Ted Mansell, Dr. Nusrat Yussouf, Mike Buban, Kent Knopfmeier, Terra Thompson, and Chris Kerr shaped many of the ideas found within this volume and solidified my understanding of convective-scale modeling and data assimilation. In addition, I'm indebted to Dr. Jack Kain, my M.S. adviser, who provided me with a strong foundation to excel as a Ph. D. student.

I've spent several periods of time at NCAR during the summers of 2011, 2012, and 2013 collaborating and picking the brains of scientists in NCAR's data assimilation section. During the summer of 2011, I was a participant in the NCAR Graduate Student Visitor Program, where much of the work in Chapter 3 initially developed and matured. Further, the experiments in Chapter 4 make use of data generously provided by Glen Romine. Stimulating discussions with Chris Snyder, Jeff Anderson, Glen Romine, David Dowell, Morris Weisman, and Craig Schwartz served

to refine the ideas and focus of the present work. Further, the NCAR/DART team, especially Nancy Collins and Tim Hoar, offered valuable technical assistance with the DART software package that facilitated the current research, and were always quick to respond to e-mail support requests.

Finally, I thank my family and the many friends I've met at OU for providing an endless stream of support and laughter to keep me grounded and sane during the last 7 years as a graduate student.

Contents

List Of Tables

List Of Figures

Abstract

Convective-scale observing system simulation experiments (OSSEs) and real-data experiments were performed to study the impact of radar and surface observations on analyses and forecasts of convective systems using the Weather Research and Forecasting (WRF) model with an ensemble Kalman filter (EnKF). The OSSEs were performed to assess the impact of covariance localization of radar data on the analyses of a developing convective system. Increasing the horizontal localization and decreasing the vertical localization produced analyses with the smallest RMSE for most of the state variables. The convective mode of the analyzed system also had an impact on the localization results. During cell mergers, larger horizontal localization improved the results. Prior state correlations between the observations and state variables were used to construct reverse cumulative density functions (RCDFs) to identify the correlation length scales for various observation-state pairs. The OSSE with the smallest RMSE employed localization cutoff values that were similar to the horizontal and vertical length scales of the prior state correlations, especially for observation-state correlations above 0.6. Vertical correlations were restricted to state points closer to the observations than in the horizontal, as determined by the RCDFs. Further, the microphysical state variables were correlated with the reflectivity observations on smaller scales than the three-dimensional wind field and radial velocity observations.

As a complement to the OSSEs, the WRF model and the EnKF were again employed to produce analyses and forecasts for the 29 May 2012 convective episode. This event produced very large hail (>4" diameter), 80 mph wind gusts, and a brief tornado, near and within the OKC metropolitan area, with estimated losses totaling 500 million dollars. Surface data, including data from surface mesoscale networks (i.e. mesonets), were assimilated at 5-minute intervals between 18 UTC and 21 UTC. Both surface and WSR-88D data were assimilated at 5-minute intervals between 21

UTC and 23 UTC, following convection initiation (CI). Several 50-member, 6-hour, ensemble forecasts were produced between 18 UTC and 23 UTC.

The frequent assimilation of surface data, especially the use of mesonet data, improved the forecast of CI timing and placement within the domain, especially for convection developing along a surface dry line. Surface data assimilation reduced a surface moisture bias that was present due to model error. Experiments where mesonet data were withheld, or where surface data were assimilated less frequently, produced less accurate forecasts of CI and possessed larger surface moisture errors. The improved surface state at 21 UTC also led to changes in the forecast convective mode after 00 UTC. The ability of sub-hourly assimilation of mesonet data to improve forecasts of CI has not been previously documented.

After two hours of both radar and surface data assimilation, the 23 UTC ensemble forecast was able to capture much of the observed convective evolution, including the tracks of several long-lived supercells. Surface data assimilation played a significant role in producing the successful forecasts. Forecasts from an experiment that assimilated only radar data contained several large errors due to a poor representation of the mesoscale environment. Some parts of the forecast were especially sensitive to the assimilation of reflectivity observations and the vertical localization of those observations. Finally, using innovation statistics, several reflectivity biases were identified in the analyses due to errors in the microphysics parameterization, the reflectivity forward operator, and biases in the environmental wind profile. Together, the OSSEs and real-data experiments herein provide a benchmark for the accuracy and ability of EnKF systems to produce analyses for more complex convective events.

Chapter 1

Motivation

Initializing numerical models with convective-scale information is essential for the accurate prediction of thunderstorms (Lilly 1990). The ensemble Kalman filter (EnKF) is one such technique that has proved itself as a capable and natural method for convective-scale data assimilation (DA), given its use of an ensemble of states to derive the needed forecast error statistics (Evensen 1994; Snyder and Zhang 2003). Unfortunately, routinely available observations that provide information on in-storm convective-scale structures are limited to Doppler radial velocity and reflectivity, yet these are indirectly, and non-linearly, related to the set of state variables needed for model initialization. Non-linearities violate several key assumptions underlying DA techniques (Kalnay 2002), but nevertheless, radar data has been used with success to generate convective-scale initial conditions (Sun 2005).

While past studies have assessed the accuracy of short-term forecasts of convective storms initialized with Doppler radar observations using an EnKF system, these cases have primarily been isolated convective events within small domains (e.g. supercells; Dowell and Wicker 2009; Dowell et al. 2011; Dawson et al. 2012, among others). The use of small domains and short-forecast lengths has led to the majority of convective-scale DA studies initializing convection within a horizontally homogeneous environment (a thorough discussion of this practice is provided in Dawson et al. 2012). Forecasts of larger-scale convective systems (e.g. mesoscale convective systems; Wheatley and Stensrud 2010; Snook et al. 2012) or using more regional domains containing a realistic variety of convective structures have received less attention. However, since the evolution of deep convection is affected by interactions between

the mesoscale and convective-scale, a better depiction of the time-varying, spatially heterogeneous mesoscale environment should improve storm-scale forecasts (Stensrud and Gao 2010; Stensrud et al. 2013).

The sensitivity of the combined mesoscale and storm-scale analyses and forecasts to observational data frequency and various components of the DA system (e.g. localization) need to be systematically examined for cases with a variety of convective structures and modes embedded within a complex mesoscale environment. Forecasts of these types of complex events will also need to predict convection initiation (CI) accurately, which represents a significant challenge (Kain et al. 2013).

Building on the successes and limitations of previous work, this dissertation explores a variety of issues that could lead to further gains in convective-scale forecast skill using an EnKF DA system, with a particular emphasis on forecasts of more complex convective events (e.g. events composed of cell mergers, upscale growth, splitting storms, etc.). These issues include: 1) the choice of covariance localization for radar data from the WSR-88D network, 2) the impact of sub-hourly surface mesonet DA on mesoscale analysis quality and CI forecast skill, 3) the relative contributions of assimilating surface mesonet vs. WSR-88D data on convective-scale forecast skill, and 4) model and forward operator errors that reduce the effectiveness of assimilating surface and radar datasets. Studies assimilating surface mesonet data and radar data together have been limited (e.g. Schenkman et al. 2011a). Keeping in mind the shortcomings of each approach, both observing system simulation experiments (OSSEs) and real-data experiments are used as tools to gain a more complete understanding of the issues noted above.

Chapter 2

Background

This chapter will focus on background material relevant to the experiments performed in the following chapters. A brief review of the sources of forecast errors and ensemble forecasting methods are provided in Section 2.1. In Section 2.2, data assimilation methods will be summarized, while Sections 2.3, 2.4, and 2.5 will summarize studies that have employed these data assimilation techniques on the convective and mesoscales, using radar and surface observations.

2.1 NWP errors and predictability

Creating an accurate three-dimensional estimate of the current atmospheric state is paramount to producing accurate weather forecasts of the future state. Lorenz (1963), using a low-order model of convection, observed that a slight perturbation to the initial conditions, originating from rounding error, would lead to deviations in the solution to the equations that would amplify with time. Lorenz suggested this instability would impose limits on atmospheric predictability, since it is impossible to completely describe the current state of the atmosphere (even in the limit of an infinitely dense, perfect, observing network, there remain scales that cannot be observed with complete accuracy). (Lorenz 1969) quantified the effects of these unobserved scales as they feedback to the larger scales over time, deteriorating the forecast; this cascade of errors imposes an *intrinsic* predictability limit on a numerical weather prediction (NWP) forecast. In addition to the intrinsic predictability limitations, observations and the methods used to produce the initial conditions both

possess errors; these types of initial conditions errors impose a *practical* predictability limit on NWP forecasts Melhauser and Zhang (2012).

Additional practical predictability limitations are imposed by NWP model error. NWP models are composed of partial differential equations that describe the fundamental atmospheric processes and their time evolution (e.g. advection, thermodynamics). These equations are discretized onto a three-dimensional grid, and are integrated in time to produce predictions of the future state, given a set of initial and boundary conditions. Using the discretized approximation to the equation set introduces error, as scales below a given threshold are not represented, but is needed since no analytical solution to the equations is known to exist. The discretization process leads to physical processes that are not represented on the model grid. These physical processes need to be parameterized to incorporate their effects on the resolved scales (e.g. boundary layer mixing, precipitation production), and this parameterization leads to further errors.

NWP accuracy is hindered by the magnitude of these initial condition errors and model errors; reducing one or both of these error sources can improve forecast skill and increase predictability (Lorenz 1965). The precise sensitivity that forecasts can be flow dependent; large errors can be tolerated in certain atmospheric flow patterns, while small errors in other patterns can lead to rapid forecast degradation (Lorenz 1965; Zhang et al. 2006b). Further, errors grow more rapidly on smaller scales (convection-scale and mesoscale) than larger scales (synoptic and global scales), leading to a shorter limit of predictability on smaller time scales (e.g. Zhang et al. 2003).

Atmospheric predictability limitations have led to the development of ensemble forecasting methods (Leith 1974; Tracton and Kalnay 1993). An ensemble forecast can provide a best solution (i.e. ensemble mean) and an estimate of the atmospheric predictability (i.e. ensemble variance). In an ensemble system, a set of initial

states is generated that encompasses the range of uncertainty in the initial state and attempts to capture the most unstable modes of error growth (Toth and Kalnay 1993). Much effort has gone into developing methods that produce an appropriate range of initial atmospheric states within the ensemble (e.g. Leith 1974; Hoffman and Kalnay 1983; Toth and Kalnay 1993; Palmer et al. 1992). Given the rapid error growth rates associated with moist convection (Zhang et al. 2006b), ensembles are necessary on the convective-scale. Recent increases in computing power have permitted convective-scale ensemble forecasts that have provided skillful predictions of convective precipitation and severe weather (e.g. Kong et al. 2006; Clark et al. 2009, 2010).

Advanced data assimilation techniques, combined with ensemble forecasting, are both ways to extend practical atmospheric predictability. Discussion of contemporary data assimilation techniques follow in the next section.

2.2 Data assimilation

Data assimilation (DA; often referred to as state estimation) is used to generate initial conditions for NWP models. DA is the process of combining many sources of information about the current atmospheric state to produce an analysis. Information most commonly originates from observations and a first-guess atmospheric state, often a climatological state estimate or one generated from a NWP model forecast in a previous analysis cycle. Various DA techniques will be reviewed below, with particular emphasis on those employed in convective-scale DA studies.

2.2.1 Optimal interpolation and 3DVAR

Gandin (1963) originally derived the multivariate optimal interpolation (OI) equations that are based on statistical estimation theory. A full derivation of the OI equations can be found in various texts, including Kalnay (2002) and Lewis et al. (2006), and are stated below without derivation for brevity. The equations are cast in matrix form, with uppercase letters denoting matrices and lowercase letters denoting vectors.

$$x^a = x^b + W(y^o - H(x^b)) = x^b + Wd \tag{2.1}$$

$$W = BH^T(HBH^T + R)^{-1} \tag{2.2}$$

OI produces an *optimal* analysis x^a, by adjusting the background state, x^b, by weighted observation innovations, Wd (Equation 2.1). The adjustments are made to x^b at the locations of the observations, thus the observations y^o need to be compared to x^b using a forward model, H, that maps x^b into observation space. The weights, W, are computed to minimize the analysis error variance using 2.2. The optimal weights produce an analysis that has less uncertainty (or at the very least, the same uncertainty) than either the uncertainty associated with y^o or x^b. Computing W is

dependent on statistical information that describes the errors associated with the observations and background, these statistics are contained within the R and B matrices, respectively.

In variational approaches, including three-dimensional variational assimilation (3DVAR), a scalar cost function is formulated that is composed of terms that represent the departure between an analysis and various pieces of information (Sasaki 1970). An example cost function, similar to that used in Gao et al. (2004), is shown below:

$$J(x) = J_b + J_o + J_c \tag{2.3}$$

$$J(x) = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(H(x) - y_o)^T R^{-1}(H(x) - y_o) + J_c \tag{2.4}$$

Terms in the cost function can include the departure between the analysis and observations, J_o, background, J_b, and dynamic or smoothness constraints, J_c, (e.g. Gao et al. 2004; Ge et al. 2012; equation 2.3). J_b and J_o are weighted inversely by R and B, respectively (equation 2.4). The final analysis is x^a is chosen as the x which minimizes the cost function. This occurs where the derivative of J becomes zero.

While OI and 3DVAR produce analyses in different ways (OI computes optimal weights that minimize the analysis error variance, while 3DVAR computes an analysis that minimizes a cost function), Lorenc (1986) demonstrated the equivalency of the two techniques. Accurately representing R and B remains a significant challenge for both approaches. B is typically represented by a static matrix containing isotropic error covariance structures. Thus, the background errors are estimated once and assumed to be stationary. This neglects the forecast "errors of the day" and, as will be discussed in section 2.2.3, is insufficient for the meso- and convective-scales where error covariance structures can be flow-dependent. R is usually constructed as a diagonal matrix, assuming that observation errors are uncorrelated.

2.2.2 4DVAR

Four-dimensional variational assimilation (4DVAR) is an extension of 3DVAR. In 4DVAR, the J_o term in the cost function is amended to include observations within an assimilation window, rather than at a single time (Talagrand and Courtier 1987). The model is integrated forward to compute the discrepancies between the observations and the analysis at each observation time. 4DVAR produces an initial condition and forecast trajectory that best fits the observations within the assimilation window.

Since the forward model is used to compute the observation discrepancies in the cost function, the model is assumed to be perfect in 4DVAR. Other assumptions include the use of a tangent-linear and adjoint model to simplify the minimization of the cost function. Further, simplified physics are used within these linearized models. Using the linearized tangent-linear model can be problematic with long assimilation windows and when simulating complex physical processes (Tremolet 2004). Even with these assumptions, 4DVAR excels over 3DVAR and OI, both due to its ability to evolve the background error covariance matrix, B and to compare each observation to an analysis state valid at the time it was observed (Lorenc and Rawlins 2005). The flow-dependent background error is a significant advantage over OI and 3DVAR, which use static estimates of B. In 4DVAR, the new analysis error covariance matrix is unavailable at the end of the assimilation window, thus each 4DVAR cycle must begin with a new estimate for B.

2.2.3 The Kalman filter

The DA techniques used herein are derived from the Kalman filter (KF; Kalman 1960; Kalman and Bucy 1961). The objective of the KF and its variants is to provide both an estimate of the state and its uncertainty (in the case of 3DVAR and OI, the uncertainty is assumed to be static, in 4DVAR it is implicitly evolved, but unavailable). The states joint probability density function (pdf) describes both

of these elements, specifying the most likely value (e.g. the mean), in addition to its uncertainty (e.g. spread). Numerically evolving the pdf is an intractable task, and thus the pdf is commonly modeled as a Gaussian. If the initial state pdf is Gaussian and the dynamic model is linear, the evolution of the state distribution will remain Gaussian, with evolving mean and covariance.

The KF algorithm consists of two components: a least-squares analysis step and a propagation step using a forward model. The analysis step uses the weighted least-squares analysis estimate from OI (2.1 and 2.2) generalized to include an update to the state uncertainty estimate, rather than relying on the assumption of a static uncertainty estimate. The propagation step assumes a linear forward model to advance both the state and error covariance estimate to the next time when observations are available. The KF algorithm is recursive since it uses a first-guess analysis and error estimate that was produced during the propagation of an earlier state and error covariance estimate, rather than recomputing previous state estimates at every analysis time. Equations describing the analysis and propagation steps are provided below, in matrix notation. The matrix dimensions are provided in parenthesis, with m denoting the number of observations, and n denoting the number of state points.

Propagation step:

$$x_k^f = M x_{k-1}^a \tag{2.5}$$

$$P_k^f = M P_{k-1}^a M^T + Q \tag{2.6}$$

Analysis step:

$$x_k^a = x_k^f + K(y_k^o - H x^f) \tag{2.7}$$

$$K = P_k^f H^T (H P_k^f H^T + R)^{-1} \tag{2.8}$$

$$P_k^a = (I - KH) P_k^f \tag{2.9}$$

where the symbols k, a, f represent time, analysis, and forecast, respectively. The model state is denoted by the vector x (n x 1), y is the observation vector (y x 1), M

is the linear model matrix that advances a discretized state forward in time ($n \times n$), P is the covariance matrix for the state estimate ($n \times n$), Q is the model error covariance matrix ($n \times n$), H is the linear observation operator mapping the state to the observation locations ($m \times n$), R is the observational error covariance matrix ($m \times m$), and K is the Kalman gain ($n \times m$). Equations 2.5, 2.7, and 2.8 are the traditional OI equations, while equations 2.6 and 2.9 are the equation to propagate and update the forecast error covariance estimate. W is denoted as K and B as P_k^f in the KF equations.

Various assumptions were made in the derivation of the KF equations. The extent to which these assumptions are satisfied determines how far the analyses stray from the optimal minimum-variance solution. These assumptions are summarized below:

1. Observations are unbiased.
2. The forecast model is unbiased.
3. Observation and forecast errors are uncorrelated.
4. Observation error covariances are known.
5. Forecast model error covariances are known.
6. Forecast model is linear.
7. Forward observation operator is linear.

Obviously, many of these assumptions will not be met for the atmospheric DA problem. The atmosphere is highly non-linear, violating (6) and potentially leading to non-Gaussian error distributions. The forward operator for observations within the present work (e.g. reflectivity) is non-linear, violating (7). Well-documented biases exist in contemporary forecast models, violating (2). For these reasons and many others, the KF update is always suboptimal. These assumptions are inherited by the extensions of the KF described below.

2.2.4 Extended Kalman filter

The use of a linear model in the KF is a limitation for some applications, thus an extension was developed termed the extended Kalman filter (EKF; Jazwinski 1970). The EKF is a modified version of the KF algorithm that implements a fully non-linear model during the state propagation step, although the linearized model is still used for the propagation of the error covariance estimate. For the update step, the background observation estimate, $H(x^f)$, is computed with the nonlinear forward operator. The linearized forward operator remains in the Kalman gain and the covariance update.

2.2.5 Ensemble Kalman filter

Calculation of the error covariance within the KF or EKF is costly for high-dimensional problems, such as those encountered in atmospheric applications. For example, the covariance evolution requires $2 * n$ applications of the model, where n is the number of degrees of freedom of the model (i.e. size of the model). Instead of evolving the entire error covariance, a *Monte Carlo* simplication to the EKF was proposed by Evensen (1994) for oceanographic applications, and subsequently applied to the atmospheric state estimation problem by Houtekamer and Mitchell (1998). This ensemble Kalman filter (EnKF) uses the fully nonlinear forward model to advance an ensemble of states that consists of a random sample of the time-evolving pdf. This ensemble propagation step replaces the covariance propagation step in 2.6, simplifying the computation; the covariance is computed using the ensemble at each assimilation interval.

Two types of EnKF update algorithms exist: a stochastic and a deterministic algorithm. In the stochastic algorithm, the observations dataset is perturbed with random noise prior to assimilation into each ensemble member (Houtekamer and Mitchell 1998; Hamill and Snyder 2002). Thus, each member assimilates a different set of observations. In the deterministic algorithm, the observations, without

perturbations, are used to first update the ensemble mean, then the ensemble is updated in a manner that produces the same analysis-error covariance as the KF. Various deterministic algorithms exist; the differences between them are summarized by Tippett et al. (2003). The ensemble adjustment Kalman filter, a deterministic algorithm, is used in this work and is described in the following section. Most recent studies use deterministic algorithms due to the additional source of sampling error that is introduced by stochastic algorithms (Whitaker and Hamill 2002).

The EnKF has several desirable properties that should give it an advantage over 3DVAR, 4DVAR, and the EKF. First, the need to develop linearized models, both for the forward dynamics and observation operator, is removed. The cost of advancing the ensemble to obtain the covariance estimate is drastically reduced compared to the EKF. Unlike 3DVAR, the EnKF provides a fully anisotropic estimate of the background-error covariances. This permits observations to adjust the state in a manner consistent with the underlying atmospheric dynamics and patterns of error growth in the model, which are often produce complex covariance structures that are flow and time dependent (Bouttier 1994). For example, observations in the vicinity of mesoscale surface boundaries can exhibit anistoropic error structures (Fig. 2.1), and covariance structures within convection are expected to be highly anisotropic (Fig. 2.2). However, the ensemble covariance estimates are affected by both sampling and model error. Various methods exist to reduce the impact of poor covariance estimates, including localization and inflation; a more in-depth discussion of these techniques will be included in Part I and Part II. Finally, the sample of the forecast pdf provided by the EnKF is a natural starting point for ensemble forecasting, thus the EnKF unifies the goals of data assimilation and ensemble forecasting.

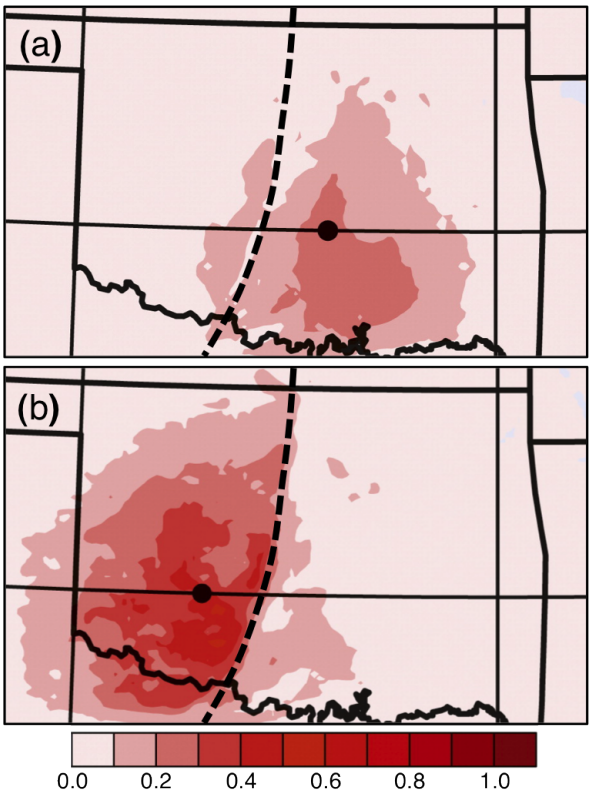

Figure 2.1: Correlation of 2-m temperature field with 2-m temperature observation located at black dot. Adapted from Knopfmeier and Stensrud (2013).

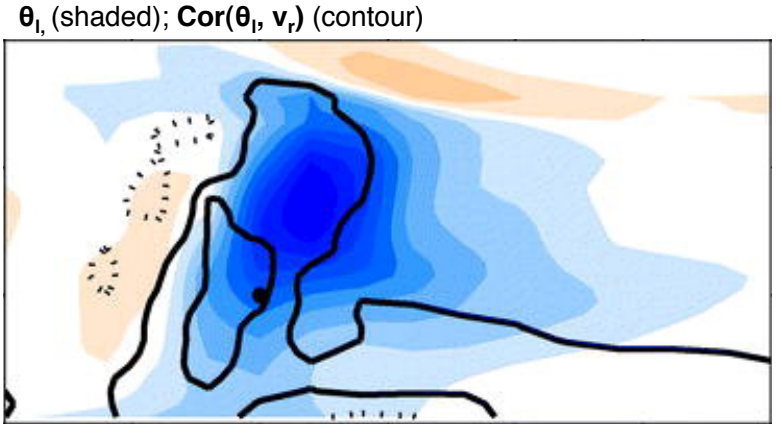

Figure 2.2: Ensemble mean liquid-water potential temperature (shaded) and correlation (solid contours, positive; dashed contours, negative) with a radial velocity observation located at the base of the updraft. Adapted from Snyder and Zhang (2003).

2.2.6 Ensemble Adjustment Kalman filter

The ensemble adjustment Kalman filter (EAKF) is a deterministic EnKF update algorithm proposed by Anderson (2001). It is mathematically equivalent to the square-root filter (EnSRF) described by Whitaker and Hamill (2002). At each data assimilation interval, the ensemble mean and members are updated using the following equations:

$$\overline{x_k^a} = \overline{x_k^f} + K(y_k^o - \overline{H(x^f)}) \tag{2.10}$$

$$\overline{x'^a} = \overline{x'^f} + \hat{K}\overline{H(x^f)} \tag{2.11}$$

$$\hat{K} = (1 + \sqrt{\frac{R}{HP^fH^T + R}})^{-1}K \tag{2.12}$$

Here, $\hat{K}$ is a reduced Kalman gain that results in smaller deviations from the ensemble mean (i.e. larger spread) compared to the full Kalman gain. Using the full Kalman gain to update each ensemble member results in excess spread reduction; in the stochastic filter the use of perturbed observations counteracts this effect.

The ensemble forecast mean and forecast error covariance in the above equations are computed with the ensemble estimate using the following equations:

$$P^fH^T = \frac{1}{N-1}\sum_{n=1}^{N}(x_n^f - \overline{x^f})\,[\,H(x_n^f) - \overline{H(x^f)}\,]^2 \tag{2.13}$$

$$HP^fH^T = \frac{1}{N-1}\sum_{n=1}^{N}[\,H(x_n^f) - \overline{H(x^f)}\,]^2 \tag{2.14}$$

Assuming the observations have independent errors uncorrelated with the background forecast, each observation can be assimilated sequentially, producing the same analysis as if all observations were assimilated simulataneously. An advantage of this method is that for each observation, HP^fH^T and R become scalars, simplifying the computation of the Kalman gain (and thus the reduced Kalman gain) and preventing the computation of a potentially large covariance matrix. This simplication is employed in the EAKF algorithm.

2.3 Convective-scale state estimation

Many of the DA methods reviewed in the previous section have been employed to estimate the three-dimensional state on the convective-scale (Sun 2005). All of these methods rely on observations from weather radar, which provide the only real source of regularly available observations within convection, at the temporal and spatial scales needed to sample convective-scale structures. The two primary observation types: reflectivity and radial velocity, are directly related to the microphysical content and wind velocity within a radar volume, and thus are indirect measurements of most quantities that are useful for diagnosing storm-scale processes or initializing a numerical model (e.g. temperature, pressure, wind, microphysical content, etc.). Thus, assimilation methods must infer the quantities of interest using additional information. DA techniques, combined with an NWP model, are used to provide these missing relationships and construct a three-dimensional convective-scale analysis.

2.3.1 Experiment design

Two types of DA experiments are presented within the dissertation: observing system simulation experiments and real-data experiments. The differences and design of each type of experiment will be discussed below, in addition to choices in the formulation of the mesoscale environment.

2.3.1.1 Observing system simulation experiments

An observing system simulation experiment (OSSE) is a type of DA experiment to completely isolate the effects of the DA system on the state estimation procedure. First, a simulation is performed using a dynamic model from a given initial state in order to create the weather event of interest. Simulated observations are extracted from this truth simulation, with errors added to each observation that are typically drawn from a Gaussian distribution with zero mean and a prescribed standard

deviation. In an OSSE, the observation locations can be completely determined by the experimenter, thus OSSEs are useful for testing the impact on analysis quality of new types and distributions of observations that may not be available at the time of the experiment. If the EnKF is chosen as the DA scheme for an OSSE, an initial ensemble needs to be constructed that samples the typical environment uncertainty. This is regularly achieved by drawing perturbations from a Gaussian and adding each realization to the truth state.

Once the truth simulation is performed and the synthetic observations are harvested, the experiment proceeds similar to a standard DA experiment. The synthetic observations are assimilated back into the model in an attempt to reconstruct the truth state. This approach assumes a perfect numerical model, although an imperfect model experiment can be designed by using a model with different characteristics (e.g. resolution, domain size, physics parameterization) during the assimilation step. The primary advantage of conducting an OSSE is that a direct comparison to the true state can be made. That is, the accuracy of the analyses due to observing system or DA changes can be isolated and quantitatively assessed.

2.3.1.2 Real-data experiments

Real-data experiments assimilate observations to estimate the true state of the atmosphere. In real-data assimilation experiments, the accuracy of the state estimate is determined through comparison with observations, rather than the truth state in OSSEs. This gives a restricted assessment of analysis accuracy, given that observations are far less plentiful than model grid points, and model error often has a significant role. Analysis quality is best assessed with observations that are not assimilated, although independent observation datasets are typically hard to obtain, given the competing desire to assimilate as many observations as possible. In these

cases, observations are compared to short-term forecasts to assess analysis accuracy, given the relationship between analysis and forecast errors.

2.3.1.3 Choice of mesoscale environment

In many EnKF OSSEs, and in some real-data experiments, that study convective-scale phenomena, a horizontally homogeneous environment is used within each ensemble member (e.g. Aksoy et al. 2009; Dawson et al. 2012). This approach permits a cleaner interpretation of the impact of the environment on the observed storm evolution through the use of sensitivity studies. Further, the interplay between the convection and its environment can be assessed, since a homogeneous environment is steady-state during the course of a simulation. Yet, to produce accurate forecasts of convection, convection must develop and evolve within a realistic convective environment. For example, forecasts of the 4 May 2007 Greensburg, Kansas, tornadic supercell were substantially improved when using a realistic inhomogeneous mesoscale environment in the experiments of Stensrud and Gao (2010). For convection that is associated with surface boundaries or due to synoptic-scale forcing, an inhomogeneous environment is vital. In many studies of isolated convection (e.g. a single supercell) over a short-time period, using a homogeneous environment is better justified, since the environment may not change substantially during the life cycle of the storm.

2.3.2 Previous convective-scale DA studies

2.3.2.1 Retrieval algorithms using Doppler velocity

Early approaches to the convective-scale state estimation problem used algorithms that combined radar observations with physical laws, such as the mass continuity equation, to generate three-dimensional analyses of convection (Sun 2005). Since radial velocity is related to the wind field through a simple geometric relationship, the three-dimensional wind field can be diagnosed using radial velocity observations

from at least two independent radars, combined with the mass conservation equation (so called dual-Doppler techniques, as first described by Armijo (1969), and later used by Ray et al. (1975), Brandes (1977), and a lengthy list of other studies). The vertical velocity field is prone to large errors due to biased estimates of horizontal divergence that contaminates retrievals (O'Brien 1970; Gao et al. 1999). Nevertheless, many studies continue to rely on dual-Doppler techniques to provide three-dimensional wind analyses within convection (e.g. Wurman et al. 2007; Markowski et al. 2008).

Deducing a mass field that is consistent with the wind field is more difficult, yet methods were developed to retrieve temperature by utilizing the equations of motion and the three-dimensional wind field at multiple times (Gal-Chen 1978; Hane and Scott 1978; Hane et al. 1981; Roux 1985). Combined, the multiple-Doppler wind analyses and mass field retrieval techniques were met with some success, especially in the determination of the wind field, but also in deducing the temperature field, for example, by determining warm anomalies in thunderstorm updrafts (Brandes 1984). On the storm scale, the retrieval of the microphysical variables is also of importance (e.g. Ziegler 1985, 1988).

2.3.2.2 Assimilating weather radar observations with 3DVAR and 4DVAR

Given the deficiencies of the retrieval algorithms (see summaries in Shapiro and Mewes 1999, Gao et al. 1999, and references therein), later work implemented fully variational approaches to retrieve the wind and thermodynamic fields. In a variational approach, the constraints that compose the cost function can be weakly imposed, (i.e. approximately, in a least-squares sense), or strongly imposed (i.e. exactly satisfied). Shapiro and Mewes (1999) used a cost function with radial wind observations and mass conservation constraints as both strong and weak constraints. Gao et al. (1999) introduced a variational analysis scheme with a cost function that included

observational, background, mass conservation, and smoothness constraints. Gao et al. (2004) incorporated a background error covariance matrix within the cost function, instead of relying on penalty coefficients as in Gao et al. (1999).

The goal of the studies discussed so far in this section was primarily to construct gridded analyses that can be used to understand the structure and salient dynamical processes acting within convection, and not to initialize NWP models. As computational power increased and radar observations became abundant over the United States with the deployment of the WSR-88D network, the explicit prediction of convection became a reality (Lilly 1990). Lin et al. (1993) initialized a cloud model with retrieved dual-Doppler winds, thermodynamic, and microphysical fields to predict deep convection. The limited availability of dual-Doppler WSR-88D observations are not widely available, motivated studies that retrieved the three-dimensional wind field from a time-series of single-Doppler data (e.g. Shapiro et al. 1995; Gao et al. 2001). Single-Doppler studies, when combined with thermodynamic retrievals, have also been used in an attempt to initialize models for features such as dry microbursts (Crook and Tuttle 1994), and when provided with an initial moisture field, to initialize convection (e.g. Weygandt et al. 2002).

Using the results of many of these previous studies (e.g. Shapiro et al. 1995; Gao et al. 2001), a DA system was developed within the Advanced Regional Prediction System (ARPS) model (Xue et al. 1995, 2000, 2001) for the explicit prediction of convection. An early version of this system, the ARPS Data Analysis System (ADAS; Brewster 1996), used a successive corrections scheme Bratseth (1986) to assimilate Doppler radial velocity observations. In addition, ADAS incorporated a cloud analysis scheme (Albers et al. 1996; Zhang et al. 1998) to adjust the moisture and latent heating profiles using reflectivity observations. Using ADAS, Xue et al. (2003) assimilated WSR-88D radial wind and reflectivity observations to initialize a forecast of the 28 March 2000 tornadic supercell event that occurred in Fort Worth,

TX. Dawson and Xue (2006) performed similar work, but for a severe MCS event. Hu et al. (2006a,b) generated improved forecasts of this event using the Gao et al. 2004 3DVAR analysis scheme and an enhanced cloud analysis scheme (Brewster 2002). This DA system has been the foundation for initializing other post-event experiments (e.g. Hu and Xue 2007; Schenkman et al. 2011b; Xue et al. 2013) and real-time forecasts generated during the NOAA Hazardous Weather Testbed Spring Experiments (e.g. Xue et al. 2008). Other 3DVAR-based systems, such as the WRF-3DVAR system, have also been developed to initialize convective-scale forecasts (e.g. Barker et al. 2004; Xiao and Sun 2007; Zhao et al. 2008). These forecasts can reduce the model spin-up period and improve precipitation forecasts in the 0-12 hour forecast period (Kain et al. 2010; Sun et al. 2012).

The 4DVAR technique is well suited to the objective of model initialization due to its use of a dynamical model. As discussed earlier, 4DVAR produces a model state by fitting the state to a time series of observations using both the dynamical model and its adjoint. This retrieves all unobserved fields simultaneously, producing a dynamically consistent three-dimensional state, including wind, temperature, pressure, and microphysical fields that can be naturally used to initialize a numerical model. The retrieval techniques discussed in this section construct a wind field first, and use it to retrieve thermodynamic fields. Sun et al. (1991) and Sun and Crook (1994) applied 4DVAR to develop a variational Doppler radar assimilation system (VDRAS) to retrieve the wind and thermodynamic fields using single-Doppler velocity data. Sun and Crook (1997) and Sun and Crook (1998), used VDRAS to produce analyses of moist convection, both in an OSSE and with real observations. Wu et al. (2000) furthered these early results. More recently, a 4DVAR system has been implemented with the WRF model (Wang et al. 2013). In 4DVAR, the development of the adjoint model, especially with the microphysical parameterizations needed for

deep convection, is both costly to develop and computationally expensive, thus other approaches have been sought.

2.3.2.3 Assimilating weather radar observations with the EnKF

The EnKF has been proposed as an alternate assimilation strategy to initialize the three-dimensional state associated with convection. An extensive amount of recent literature has devoted attention to assimilating weather radar observations, both Doppler radial velocity and reflectivity, using the EnKF, beginning with the initial OSSE study of Snyder and Zhang (2003). Others have used OSSEs to examine various fundamental challenges with radar DA using the EnKF (e.g. Zhang et al. 2004; Tong and Xue 2005; Caya et al. 2005; Xue et al. 2006; Jung et al. 2008; Xu et al. 2008; Lu and Xu 2009; Thompson et al. 2012). Alongside these developments, others have experimented with the assimilation of research or operational radar datasets (e.g. Dowell et al. 2004; Aksoy et al. 2009; Dowell and Wicker 2009; Dowell et al. 2011; Snook et al. 2011; Dawson et al. 2012; Jung et al. 2012; Tanamachi et al. 2013; Yussouf et al. 2013).

These studies have assimilated Doppler radial velocity and/or reflectivity observations. While radial velocity is directly related to the three-dimensional wind field, radar reflectivity depends on the specifics of the hydrometeor distribution, including number, size, phase, and orientation. Predictions of the evolution of hydrometeor distributions are parameterized, thus described by analytical functions that evolve using uncertain estimates of precipitation process tendencies, leading to significant model error. Additional errors exist due to uncertainties in the computation of reflectivity from the microphysical variables in the model state (i.e. forward operator errors). For example, the increase in the sampling volume of the radar beam with range is not usually taken into account in the forward operator. Further, the relationship between reflectivity and the model state is non-linear,

violating the conditions for optimality in the EnKF and leading to a suboptimal update. Finally, reflectivity observation errors can be large due to miscalibration of the radar system and signal attenuation. Together, these errors sources present a formidable challenge to assimilating radar reflectivity.

To isolate the effects of reflectivity assimilation on storm-scale analyses, Dowell et al. (2011) conducted sensitivity tests using radar observations of an observed supercell. While reflectivity observations helped to accelerate the spin-up of convection in the analyses (especially at far distances from the radar), biases contributed to large errors in within the storm analyses. They attributed these errors to be predominately from two sources: 1) the reflectivity forward operator, especially in the computation of reflectivity contributions from ice and mixed-phase precipitation, and 2) microphysical parameterization scheme reflectivity bias. Errors from the second source were noted within the downshear anvil at low-levels, hypothesized to be due to errors in the models prediction of rainwater mixing ratio within the supercell forward-flank. Persistent, unnecessary, adjustments to temperature, via cross-correlations between reflectivity and temperature, played a role in producing forward-flank cold pools that were too intense. Limiting the DA update to the hydrometeor fields ameliorated this bias, at the expense of a longer spin-up period. These biases are common in single-moment microphysics schemes of the type used in Dowell et al. (2011), and may be reduced by using schemes that predict more moments of the hydrometeor distribution, such as number concentration or reflectivity (Dawson et al. 2010).

2.3.2.4 Ensemble forecasts of convection using the EnKF

Only a small number of studies have analyzed short-term forecasts of convective systems with initial states generated by the EnKF. Aksoy et al. (2010) used an EnKF ensemble to initialize a 30-minute forecast of three different convective events (a

supercell, multi-cell, and squall line). Their experimental design used a homogeneous environment without parameterization of boundary layer or surface processes. They surmised that these simplifications led to the some of the ensemble forecast errors, including inaccurate predictions of storm maintenance due to environments that did not support convection. They noted that these results demonstrated the importance of incorporating mesoscale information, both in terms of the environment and its uncertainty, for successful convective-scale data assimilation and forecasting.

Even in a relatively spatially homogeneous environment, such as that associated with an isolated supercell, important temporal changes in the wind and temperature profiles can have marked effects on short-term forecasts. These changes are often dramatic over the lifecycle of a convective storm, as the loss of daytime heating causes changes to the low-level temperature and wind profiles (e.g. low-level jet) that can modulate storm motion, intensity, and severity. Dawson et al. (2012) produced 1-hour forecasts of a tornadic, isolated supercell, in an environment characterized by a rapidly strengthening low-level jet. Various low-level wind profiles, taken over the course of a one-hour period, were used to initialize a homogeneous environment. Forecasts of the area impacted by significant low-level rotation within the supercell were quite sensitive to the chosen low-level wind profile, even over this short time period. They concluded by noting that initializing forecasts in these types of environments is problematic.

While Aksoy et al. (2010) and Dawson et al. (2012) produced 1-hr ensemble forecasts from an EnKF system, Snook et al. (2012) assimilated WSR-88D and CASA radar data using the EnKF and produced 3-hr ensemble forecasts for an MCS containing tornadic mesovortices. Snook et al. (2012) focused primarily on sensitivities of the forecasts of mesovortices to microphysical parameterization and the assimilation of high-resolution CASA radar data.

2.3.2.5 Some challenges for successful radar DA

Much of the existing literature has focused on isolated convective events, oftentimes a single, confined area of convection such as an isolated supercell. Radar DA techniques, especially those involving the ensemble Kalman filter (EnKF) have matured by studying these events almost exclusively. For the more complex convective event discussed herein, challenges remain with the successful assimilation of radar data, including both reflectivity and radial velocity.

One successful assimilation strategy that has been used with previous studies is to use additive noise to promote the rapid spin-up of convection and maintain ensemble spread (Dowell and Wicker 2009). To achieve the former objective, the magnitudes of noise added to the analyses are often the largest at the beginning of the assimilation period, and then are reduced to a smaller magnitude later in the period to achieve the latter goal. Even though the additive noise procedure can exacerbate temperature errors within the surface cold pool and is largely an ad hoc technique, it is regarded as having an overall positive benefit on the convective analyses (Dowell et al. 2011) and is used extensively in storm-scale EnKF radar DA studies.

A similar strategy is used with the assimilation of reflectivity observations; these observations promote the rapid spin-up of convection, but are often detrimental to the analyses after a spin-up period. Thus, reflectivity assimilation is often ended after an initial predetermined time period, with later assimilation cycles assimilating only radial velocity and clear-air observations. It is unclear how these two assimilation strategies can be generalized to a more complex convective event such as the one studied herein. In this case, multiple areas of convection are developing at different times through the assimilation period. Turning off reflectivity assimilation or reducing additive noise at a prescribed time over the entire domain would potentially hinder the development of later convection, and be detrimental to convection that has already spun-up prior to the cut-off time.

A secondary challenge is the use of clear-air radar observations (i.e. a subset of reflectivity observations less than a predetermined threshold, usually around 10 dBZ) that serve to prevent spurious convection during the assimilation period. In general, these observations have been assimilated following the initial development of convection. This is a natural consequence of initializing experiments near or slightly following the time of CI, when reflectivity and radial velocity observations first become available. In the present case, CI is observed in different times and regions within the domain. Using the traditional approach of assimilating clear-air observations when precipitation first develops within the domain may adjust the environment toward a state that is unfavorable for convection in areas where convection may later develop. In these regions, the adjustments that take place may have to be reversed when reflectivity and radial velocity observations within precipitation become available. The ramifications of this process, and thus the most effective use of clear-air observations, remain largely unknown.

2.4 Mesoscale state estimation at the surface

Observations near the surface (e.g. temperature, dew point temperature, horizontal wind components) have the potential to be utilized in DA systems to provide more accurate estimates of the temperature and moisture profiles within the planetary boundary layer (PBL), as well as the placement and strength of near surface boundaries (e.g. fronts, dry lines). These surface boundaries often serve to focus the development of deep convection. Once convection has developed, the resulting mode, intensity, and coverage of convection is strongly controlled by the temperature and moisture profile within the PBL. Thus it is a desirable goal to improve estimates of the structure, placement, and horizontal and vertical circulations associated with these boundaries, in addition to the representations of the PBL. These improvements should lead to more accurate forecasts of CI and its subsequent evolution. Efforts

to assimilate surface observations in numerical models, with specific emphasis on improving forecasts of convection, are summarized in this section.

2.4.1 Assimilating surface observations into NWP models

Many surface observation networks exist that provide high-resolution temporal and spatial data (e.g. Oklahoma Mesonet, West Texas Mesonet). These networks provide a rich data source that can be assimilated into numerical forecast models. So far, these data have been underutilized operationally due to challenges with their assimilation (Pu et al. 2013). For example, mismatches often occur between the model and surface observations in areas of terrain, where, due to model resolution, the model surface may not be at the same height as the surface observation height. Also, surface observations are not always coupled to the overlying atmosphere, thus their impact on the column above the observation location is difficult to prescribe and control. Further, the parameterized PBL and surface layers are often biased in environments conducive to the development of deep convection (Coniglio 2013), leading to suboptimal, and potentially detrimental, assimilation performance.

Even in light of these challenges, recent studies have demonstrated some success in assimilating surface observations on the mesoscale using the EnKF. Using perfect and imperfect model experiments, Zhang et al. (2006a) and Meng and Zhang (2007) showed that assimilating surface observations every 3-hours improved analyses and forecasts of a poorly forecast snowstorm. Meng and Zhang (2008a) extended the previous two studies by assimilating real surface and rawinsonde data during a mesoscale convective vortex event. In Meng and Zhang (2008a), surface observations were assimilated every 6-hours, and reduced forecast error, although surface observations had a much smaller impact compared to sounding and profiler data, and thus were not assimilated in a follow-up study (Meng and Zhang 2008b). In these studies, the models horizontal grid spacing was 30 km. Surface observations

were placed every 60 km (in Zhang et al. 2006a and Meng and Zhang 2007) or where routine[1] surface observations were available (in Meng and Zhang 2008a). Other studies have investigated the use of the EnKF for assimilating surface observations over complex terrain (Pu et al. 2013; Ancell et al. 2011). Again, routine observations were used in these studies on model grids ranging from 12-km (Ancell et al. 2011) to 27-km (Pu et al. 2013).

Other studies have focused more closely on how to effective assimilate surface observations and their ability to identify and reduce model errors within a parameterized planetary boundary layer (PPBL). Using a perfect and imperfect single-column model and simulated observations, Hacker and Snyder (2005) demonstrated that assimilating surface observations, using the EnKF, constrained the entire profile within the PPBL, especially when the surface was deeply coupled with the atmosphere (e.g. during the late afternoon when the PPBL was well-mixed). The height of the PPBL was also improved when surface observations were assimilated. Hacker and Rostkier-Edelstein (2007) arrived at similar conclusions using real observations over Oklahoma.

2.4.2 Assimilating surface observations: Convective weather applications

Several studies have focused more directly on the benefits of surface observations on mesoscale predictions of severe weather. Fujita et al. (2007) exclusively assimilated surface observations each hour onto a 30-km model grid during the 12 UTC - 18 UTC period preceding two severe weather events. An 18-hour ensemble forecast was initialized at the end of each assimilation period. The authors noted improvements in the placement and intensity of mesoscale boundaries such as dry lines and fronts,

[1]Routine surface observations include METAR and exclude observations from mesonetworks (e.g. Oklahoma mesonet).

in PBL depth and structure, and in resulting placement of precipitation both in the analyses and throughout the ensemble forecast period. Stensrud et al. (2009b) demonstrated improved analyses of two mesoscale convective system (MCS) events by assimilating hourly surface temperature, moisture, and wind observations. Wheatley and Stensrud (2010) expanded on the Stensrud et al. (2009b) study by isolating the potential positive benefit of assimilating total surface pressure (via altimeter) observations on analyses of MCS cold pools, rather than the assimilation of 1-hour surface pressure tendency. Wheatley et al. (2012) assimilated surface observations, along with radiosonde, marine, and aircraft observations, each hour during a 5-hour period and produced improved short-term forecasts of severe weather parameters in 0-12-hr ensemble forecasts. The specific impact of surface observations in Wheatley et al. (2012) is unknown. Routinely available surface observations were assimilated in all of the previous studies mentioned so far. In addition to routine observations, Knopfmeier and Stensrud (2013) assimilated mesonet observations every hour for two two-week periods and, through data denial experiments, determined that the mesonet observations did not have a significant impact on the resulting analyses. They suggest that a clearer benefit may result by using more frequent assimilation cycles.

Several studies have noted that assimilating surface observations might be a boon to forecasts of CI. For example, while discussing the beneficial impact of surface data on boundary layer profiles in their idealized results, Hacker and Snyder (2005) commented that the *"extension of these results to real modeling systems could improve forecasts of pollutant transport and storm initiation."* Yet, few studies have focused specifically on the effects of surface DA on short-term forecasts of CI and convective evolution. Previous work has focused primarily on improved representation of the environments using grid spacings of 12 km or greater. At these grid lengths, convective parameterization is required, thus the benefits to convection initiation and evolution cannot be explicitly addressed, but only inferred through the convective

precipitation output from the parameterization scheme. For example, Ancell (2012), surface observations were assimilated every 3-hours on a 4-km grid, while Knopfmeier and Stensrud (2013) used a 5-km grid, but neither of these studies focused on convective development (in Ancell 2012 the model domain covered only portions of the Pacific Northwest). Further, with grid spacings ranging from 12 km to 30 km, the model grids only are able to resolve features with spatial scales within the meso-alpha scale (100-1000 km). Smaller scale features that often play a role in initiating convection (e.g. boundary layer rolls, smaller outflow boundaries) are left unresolved.

Observations gathered during the International H_2O project (IHOP; Weckwerth et al. 2004) spurred a collection of studies focused on understanding the key processes involved in CI (e.g. Weckwerth and Parsons 2006) and predicting CI using NWP models with convection-permitting grid spacing (e.g. Xue and Martin (2006a,b)). While most of CI modeling studies assimilated surface observations, including special observation sets collected during IHOP, their goals differed. Xue and Martin (2006a), hereafter XM06a, focused on the skill of short-term model forecasts (3-hr lead time) at predicting CI locations for the 24 May 2002 IHOP case. XM06a used an initial analysis produced by assimilating routine and special IHOP upper-air and surface observations with ADAS (Brewster 1996). Although observations were assimilated at only two times (12 UTC and 18 UTC), the deterministic forecast produced a fairly accurate prediction of CI and the general evolution of the convective line. Other studies focused on identifying the physical mechanisms leading to CI in simulations similar to those produced in XM06a (Xue and Martin 2006b), hereafter XM06b, and Wang and Xue (2012). None of these studies systematically investigated the impact of the various observations types or assimilation strategies on the resulting analyses and forecasts.

Liu and Xue (2008), hereafter LX08, followed a similar approach to XM06a and XM06b to study the IHOP case on 12 June 2012, but performed an additional set

of sensitivity experiments to examine how varying DA intervals, correlation length scales, and nonstandard observations types affect predictions of CI and evolution. The timing and placement of multiple areas of CI in the various experiments were sensitive to these choices. An experiment with 3-hourly assimilation cycles during a 6-hour period provided a better overall forecast of CI timing, while an experiment with hourly assimilation provided a better forecast of convective evolution. Yet, the hourly assimilation experiments improved the forecast of the low-level wind field, compared to a 3-hour and 6-hour assimilation period. LX08 suggested that the delay in CI observed in some of their sensitivity experiments was due, in part, to the weakening of resolved low-level horizontal convergence after each assimilation cycle, since the assimilated surface observations have insufficient spatial resolution to provide convective-scale information. Finally, an experiment with a larger vertical correlation length scale resulted in a deeper, longer-lived cold pool, leading to increased convergence along the cold pool gust front. This improved forecasts of CI initiated by the gust front, but hindered CI and evolution in other regions.

To the best of this authors knowledge, LX08 is the only study that has explored how surface observation datasets and assimilation strategies affect convective-scale analyses and forecasts of CI and evolution. Thus, while some of the goals of the present work are shared with the LX08 study, substantial differences exist. First, ADAS, as used in LX08, implements the Bratseth (1986) successive correction method to produce the analyses. This system is now largely obsolete in the NWP community, replaced with more advanced DA algorithms, including state-of-the-art schemes that provide flow-dependent estimates of the forecast errors (e.g. EnKF). LX08 realized the need for such capabilities, stating, *"for truly optimal analysis, flow-dependent background error correlation scales have to be estimated and used."*

2.5 Combined surface and radar data assimilation

While surface observations have the potential to improve CI forecasts, their use is also important following CI, when radar observations become available. If radar observations are being assimilated in regions where the mesoscale environment is unsupportive of convection, for instance, due to errors in the placement of critical surface boundaries, then convection can fail to become established in the model and dissipate rapidly in any forecasts initialized with these analyses. Assimilating surface observations can ensure the simulated environment supports convection and improve the evolution of convection beyond the early stages of the forecast by making lasting adjustments to the mesoscale environment.

Several studies have assimilated both surface and radar data for the multi-scale prediction of convection, but most have used variational, not ensemble, approaches. For example, the ARPS ADAS and 3DVAR system have been frequently used to assimilate both observation types for the prediction of a variety of events (e.g. Dawson and Xue 2006; Hu and Xue 2007; Schenkman et al. 2011b). Since most EnKF real-data radar DA studies have used a homogeneous environment to predict isolated areas of convection (e.g. Dowell et al. 2011; Dawson et al. 2012), the utility of surface observations is limited. Surface observations would only be useful within convectively-generated features, such as the surface cold pool. The lack of routinely available surface measurements on the convective-scale has precluded the use of these observations in storm-scale DA studies that employ homogeneous environments, although recently Marquis et al. (2014) used mobile mesonet observations from a field experiment to constrain the development of the surface cold pool for a supercell event. Surface observations are more useful in a heterogeneous environment, yet few examples exist where both surface and radar data were assimilated on a convection-permitting grid. Most recent work has assimilated surface observations on a coarser grid, while assimilating radar data on a finer grid (e.g. Yussouf et al. 2013). Further

discussion of these studies and their relation to the current work is provided in Chapter 4.

Chapter 3

Part I: Observing system simulation experiments of a developing convective system

An OSSE was designed to explore the capability of EnKF data assimilation to produce analyses of a more complex convective event, such as that associated with a developing convective system that grows upscale through cell and cold pool mergers. An important aspect of the EnKF is analysis sensitivity to choices for covariance localization, which will be the focus of this chapter. The combination of the type of convective system and systematic study of localization sensitivities make the present OSSEs unique.

A summary of localization choices in convective-scale DA literature is provided in the next section. Section 3.2 describes the experimental design and section 3.3 and 3.4 examine the impacts of the localization cutoff, r, on analysis error, and the analysis increments respectively. Section 3.5 examines the correlation length scales between the observations and state, while Section 3.6 examines short-term ensemble forecasts spawned from the final EnKF analyses. Section 3.7 contains conclusions and a synthesis of the results. Much of this chapter is derived from Sobash and Stensrud (2013)[1].

3.1 Localization for radar data assimilation

Insufficient ensemble size leads to degraded background error covariance statistics due to sampling error, resulting in a sub-optimal data assimilation update that

[1]©2013, American Meteorological Society

is propagated forward by the ensemble advance. For convective-scale EnKF applications, having a large enough ensemble of model simulations (>>100 members) to mitigate sampling errors is computationally prohibitive, thus these errors are inevitable for such experiments. A ubiquitous treatment to reduce these errors is to confine the impact of an observation to nearby model grid points. Errors in the sample correlation are greatest where the true correlations are small (Fisher 1915), thus the true correlation structure is more robustly measured at grid points near the observation, where the correlations should be largest. Thus, using localization implies that distant estimates of non-zero correlations are more vulnerable to sampling error and should be ignored.

Localization is commonly implemented by weighting the observation-state covariance estimates from the ensemble using a function that becomes zero beyond a specified cutoff distance. This was first applied using a quasigeostrophic model by Houtekamer and Mitchell (1998), hereafter HM98, and further refined to utilize a Schur (elementwise) product for filtering noisy covariances. This has several benefits, including reducing computational burden by only updating a subset of model state points, and increasing the rank of the background error covariance matrix (Hamill et al. 2001).

Improper localization can introduce imbalance into the analysis, thus optimally choosing localization is essential for numerical weather prediction (Greybush et al. 2011). Flow-dependent error covariance structures vary in time and space; adaptive localization techniques that permit these structures (e.g. Bishop and Hodyss 2009a,b) could provide more benefit than the fixed weighting schemes described above, but have not yet been applied to convective-scale EnKF.

All previous convective-scale EnKF studies have implemented some manner of covariance localization (Table 3.1). When implementing localization, the two

adjustable parameters are the weighting function and the cutoff distance[2], r. Early studies used a boxcar weighting function (e.g. HM98), while later studies (e.g. Houtekamer and Mitchell 2001) have implemented a weighting function that approximates a Gaussian function (Gaspari and Cohn 1999; hereafter GC99). Often, tuning is used to select an optimal r for the weighting function, although the specific sensitivities often go undocumented or a cutoff is chosen following recommendations in previous work. Ultimately, the optimal localization cutoff depends on the ensemble size, state variable, observation type, density, location, and the dynamics of the observed system. Many of these characteristics differ among the previous studies, preventing any robust conclusions about the effects of localization on the analyses, especially those that utilize WSR-88D radar datasets.

Early studies using OSSEs chose r between 4 and 8 km, with equivalent values for the horizontal and vertical cutoffs (i.e., the region of influence around an observation is a sphere). For example, Snyder and Zhang (2003), in the first documented convective-scale perfect-model OSSE of a supercell, used $r = 4$ km for a 50-member ensemble with a boxcar weighting function. While not the focus of their work, their results showed some sensitivity to the cutoff, with large r values (>20 km) being detrimental to the analyses and little sensitivity for r values between 2 km and 6 km. Using the GC99 weighting function improves their results over the boxcar function, and thus later studies have almost exclusively used the GC99 function for localization.

Caya et al. (2005) further examined the analysis error of differing localization cutoffs. The smallest errors occurred using a value of 7.3 km for r in their 100-member OSSEs, but they noted that a smaller r (approximately 5.5 km) should be used for a 50-member ensemble. Tong and Xue (2005) also used a 100-member ensemble, but with $r = 8$ km. In a real-data study, Dowell et al. (2004) used $r = 6$ km, and

[2]Here, the cutoff distance, r, is the distance beyond which the localization weights are zero. Some literature defines the term cutoff to mean a half-width, that is, $0.5r$.

Reference	Ensemble Size	Analysis grid (km)	Radar obs spacing	Localization cutoff (r, km)
Snyder and Zhang (2003)	50	2	Model grid	H: 4; V: 4
Dowell et al. (2004)	50	2	Radar scans	H: 6; V: 6
Tong and Xue (2005)	100	2	Model grid	H: 8; V: 8
Caya et al. (2005)	100	2	Radar scans	H: 7.3; V: 7.3
Aksoy et al. (2009)	50	2	Radar scans	H: 5; V: 4
Dowell and Wicker (2009)	50	1	Radar scans	H: 6; V: 6
Dowell et al. (2011)	50	1	Radar scans	H: 6; V: 6
Dong et al. (2011)	50	2	Radar scans	H: 6; V: 6
Dawson et al. (2012)	30	1	Radar scans	H: 12; V: 6

Table 3.1: Summary of covariance localization choices in a selection of previous convective-scale EnKF studies. Radar observation spacing of "radar scans" indicates observations are assimilated at model grid points in the horizontal, but kept on the radar sweep surfaces in the vertical. "Model grid" indicates observations are located at model grid points in both the horizontal (H) and vertical directions (V).

concluded that cutoff radii between 4 and 10 km produced the best results, while r = 2 km was detrimental to the analysis. Aksoy et al. (2009) chose r = 5 km (4 km) horizontal (vertical) cutoff, and found little sensitivity to a larger horizontal cutoff radius, but noted larger sensitivity to the vertical localization cutoff. Later studies generally settled on r = 6 km (Dowell and Wicker 2009; Dowell et al. 2011; Dong et al. 2011), although Dawson et al. (2012) used a larger horizontal cutoff (r = 12 km) than all previous studies. In other real-data studies, attempts were made to extract information about multiple scales of motion from radar observations by using varying localization cutoffs for different groups of observations (Zhang et al. 2009). While this was employed to produce analyses of a hurricane using dense radar observations, the technique is also potentially useful for convective storm analyses.

In an attempt to isolate the impacts of changes to covariance localization on convective-scale EnKF analyses, perfect-model OSSEs of a developing mesoscale convective system (MCS) were conducted using a variety of cutoff distances encompassing the range of values used in previous work. MCSs are understudied in the convective-scale EnKF literature, with supercells garnering most of the attention (all the referenced studies in Table 3.1 have focused on supercells, with the exception of Aksoy et al. (2009) which included a supercell and squall-line case). In addition to testing the sensitivity of the EnKF analyses to the localization cutoff, the results herein could provide evidence of differences in localization between convective modes (e.g., isolated supercells vs. squall lines).

To capture such differences, the present OSSEs simulate the process of upscale growth from initially isolated cells into a linear squall line during a 150-min period (much longer than any previous OSSE or real-data convective-scale EnKF experiment). The main thrust of this work aims to understand the impacts of changes in r on the error of the state fields and by doing so, gaining an understanding of the role of localization in producing accurate convective-scale analyses across a spectrum

of convective modes. Further, the proper choice of localization is related to the covariance structures present within a simulated ensemble of convective storm. Little attention has been paid in the convective storms literature to understanding these structures; the analyses within this chapter relate the covariance structures to the localization choices.

3.2 Methods

3.2.1 Truth simulation

The Advanced Research Weather Research and Forecasting (WRF-ARW) model (Skamarock et al. 2008) V3.2.1 was used to create a truth simulation of a developing MCS. The model was initialized using a thermodynamic profile derived from 28 proximity soundings from strong, weakly-forced MCS events (Coniglio et al. 2006; Fig. 3.1). Coniglio et al. (2006) tested a variety of wind shears in the environmental wind profile; the wind shear used herein produced a strong, long-lived, upright-tilted squall line by the end of the simulation. This shear profile (Fig. 3.1) contains 20 m s^{-1} of 0-5 km westerly shear, sufficient for long-lived squall lines in model simulations (Weisman and Rotunno 2004). Convection was triggered with 5 +3K bubbles, with random perturbations added along the bubble edges to initiate three-dimensional motions. The horizontal grid spacing is 3 km and spanned 396 km in both horizontal directions (133 grid points), and 16 km in the vertical (40 grid points). These dimensions, larger than most EnKF convective-scale experiments, were required to retain the system within the domain during the 3-hour period. The WRF-ARW was integrated as a cloud-model with WRF 6-class single-moment (WSM6) microphysics (Hong and Lim 2006), open-boundary conditions, no terrain, surface physics, or boundary-layer physics. The Coriolis force was permitted to act on the perturbation wind, as is customary in convective-scale cloud modeling. A prognostic 1.5 order turbulence

kinetic energy closure is used to compute the horizontal and vertical eddy viscosities necessary to represent the effects of sub-grid scale turbulence.

3.2.2 Simulated radar observations

Synthetic radial velocity and reflectivity observations were extracted from the truth simulation every five minutes assuming a radar placement at the center of the domain. A severe weather WSR-88D scanning strategy composed of 14 scan elevations (between 0.5° and 19.5°) was used. The observations were extracted from each grid column within the range of the radar (230 km), at the intersection of the grid columns and the 14 scan surfaces. This technique for extracting simulated radar observations was also applied in Xue et al. (2006) and serves to reduce the number of observations and produce a gridded observation dataset. The observations were computed by trilinearly interpolating the required model state fields to the location of each simulated radar bin. Observation errors were added to each synthetic observation by performing a random draw from a Gaussian with zero mean and standard deviation of 2 m s^{-1} (2 dBZ) for radial velocity (reflectivity). Reflectivity observations of less than 10 dBZ were considered clear-air observations for the purposes of this study and the corresponding radial velocity observation at these locations was removed. All other observations are considered precipitation observations.

3.2.3 Data assimilation and OSSEs

The assimilation experiments were conducted with the EnKF implementation within the Data Assimilation Research Testbed (DART) software[3]. Specifically, the sequential parallel version (Anderson and Collins 2007) of the deterministic ensemble adjustment Kalman filter (Anderson 2001) was used for the data assimilation update. The 50-member ensemble was initialized with different horizontally homogeneous

[3]DART software available at http://www.image.ucar.edu/DAReS/DART

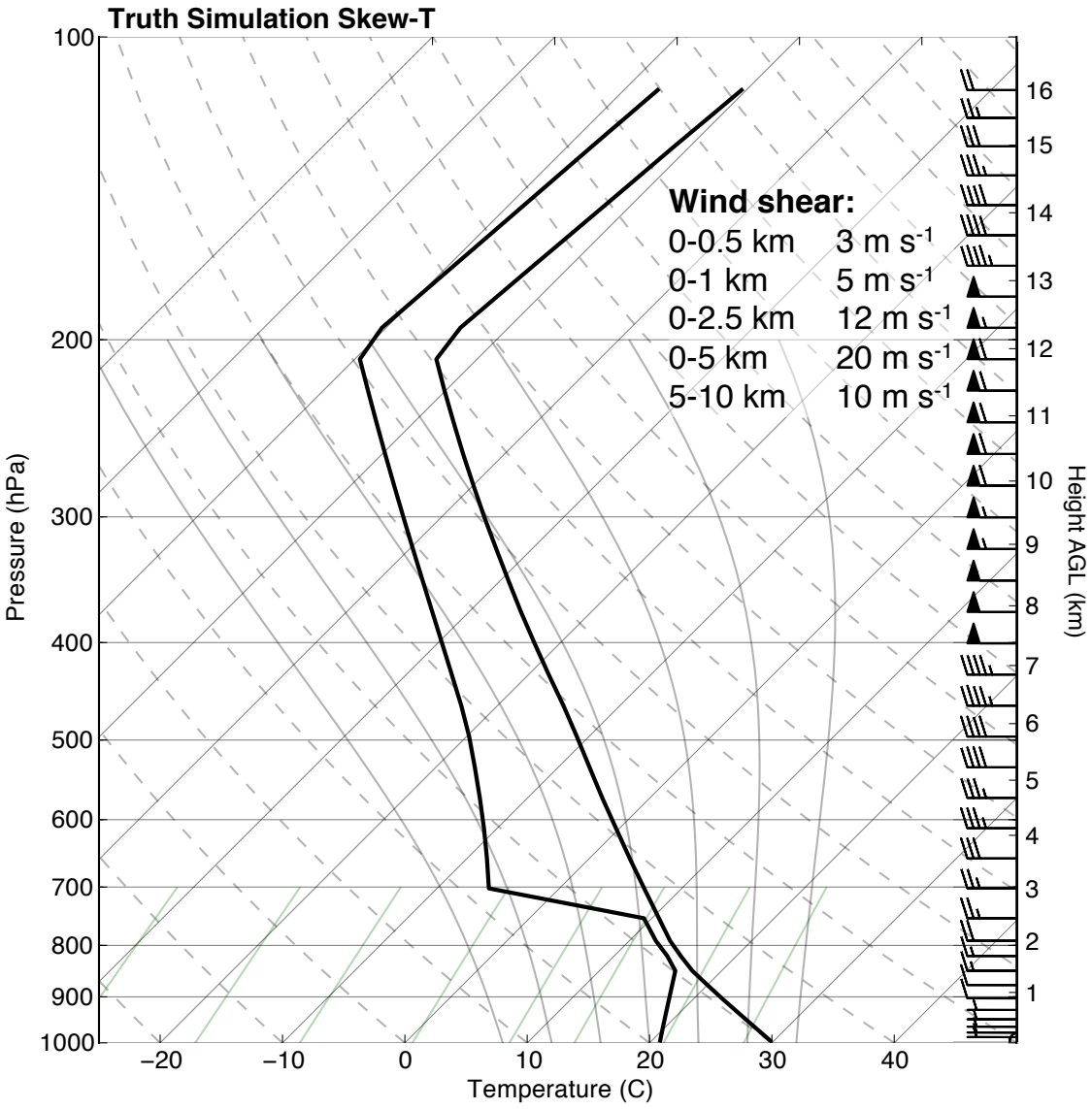

Figure 3.1: Skew-T profile used to initialize truth simulation. Taken from Coniglio et al. (2006). Sounding represents the median environmental thermodynamic profile of 28 weakly-forced MCSs. The unidirectional environmental wind profile is constructed to contain values of environmental shear that produce intense simulated squall lines.

wind profiles (identical thermodynamic profiles), each produced by adding random, uncorrelated Gaussian noise to the wind profile (both U and V) of the truth sounding below 11 km AGL. The ensemble generation technique used herein is consistent with other OSSEs and real-data experiments that use homogenous environments for initial conditions (e.g. Dowell et al. 2004). The initial ensemble was the same for each OSSE. The OSSEs were initialized 30-minutes after the beginning of the truth simulation when appreciable amounts of hydrometeors develop within the simulated storm. The synthetic observations were assimilated every 5-minutes for 150 minutes. To increase ensemble spread, additive, smoothed, perturbations were added to the wind, temperature, and dew point temperature fields where the observed (i.e. simulated) reflectivity is greater than 20 dBZ immediately prior to the ensemble advance (Dowell and Wicker 2009) throughout the length of the assimilation period. The additive noise smoothing procedure prevents any detrimental effects due to discontinuities on the edges of the convective system. The magnitudes of additive noise (Table 3.2) were chosen using values from previous work (e.g. Dowell and Wicker 2009). The additive noise magnitudes were not tuned for each experiment, yet the results herein appear to be robust. Analysis of additive noise sensitivities are found in section 3.4.3.

The OSSEs use the same model configuration as the truth simulation (i.e., they are perfect-model experiments). The experiments are identical except for the localization cutoff, r. r_h varies between 6, 12, and 18 km in three experiments to assess the impacts of the horizontal localization ($r_v = 6$ km in all three experiments), while in another set of experiments, r_v is varied between 3, 6, and 12 km while r_h is held constant at 12 km (Table 3.2). The localization choices for the clear-air observations remain constant across all experiments; those observations are thinned to 6 km in the horizontal with $r_h = r_v = 12$ km. These clear-air observations have been shown to improve convective-scale EnKF analyses by suppressing spurious convection (Aksoy et al. 2009).

Experiment	$r_{h,precip}$ (km)	$r_{v,precip}$ (km)	$r_{h,noprecip}$ and $r_{v,noprecip}$ (km)	Noise (K or m s^{-1})
LOC6V6	6	6	12	1.0
LOC12V6	12	6	12	1.0
LOC18V6	18	6	12	1.0
LOC12V3	12	3	12	1.0
LOC12V12	12	12	12	1.0
LOC12V12N0.5	12	12	12	0.5
LOC12V12N0.25	12	12	12	0.25

Table 3.2: Summary of covariance localization experiments conducted in Chapter 3. The first number in each row indicates the horizontal localization cutoff (km) and the second number indicates the vertical localization cutoff (km) for precipitation observations. Hence, LOC12V6 uses a horizontal and vertical localization cutoff of 12 km and 6 km, respectively. All experiments use the same horizontal and vertical cutoffs for clear-air (no precipitation) observations.

3.2.4 State-space diagnostics and analysis techniques

The fidelity of the analyses to the truth simulation is gauged by domain-averaged grid point based verification metrics, as is common in EnKF analysis verification (e.g. Dowell et al. 2004). In an OSSE, the state estimate can be directly compared to the true state at all state points. Herein, these state-space diagnostics (as opposed to observation-space diagnostics that are computed at the observation locations) are the primary metrics used in this study to assess the overall quality of an OSSE. Root-mean-squared error (RMSE) of the analyses and 5-min forecasts is computed, defined as,

$$RMSE^{a,f} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(x_i^t - \overline{x_i^{a,f}})^2} \tag{3.1}$$

where the subscript i is an index over all M grid points where the true state reflectivity is greater than 20 dBZ, and t, a, and f represent the true, analysis (i.e., posterior), and forecast (i.e., prior) state, respectively. The use of clear-air observations prevents the development of spurious convection outside the truth state convection (this was confirmed by visually inspecting the ensemble for areas of convection outside the truth state), justifying the use of the truth state for defining the points over which RMSE is computed. State-space RMSE is computed for the three velocity components, temperature, and the WSM6 microphysical species (water vapor, cloud water, cloud ice, rain, snow, and graupel). In addition, the average bias is computed, defined as,

$$BIAS^{a,f} = \frac{1}{M}\sum_{i=1}^{M}(x_i^t - \overline{x_i^{a,f}}) \tag{3.2}$$

This is analogous to the mean innovation in observation-space. The consistency ratio (CR), a comparison of the ensemble mean RMSE to the ensemble spread, is also computed to ensure the ensemble contains an appropriate amount of spread. This ratio is defined as,

$$CR^{a,f} = \frac{\frac{1}{M}\sum_{i=1}^{M}[\frac{1}{N-1}\sum_{n=1}^{N}(x_{i,n}^{a,f} - \overline{x_i^{a,f}})^2]}{(RMSE^{a,f})^2} \tag{3.3}$$

where the numerator is the squared ensemble spread (i.e., ensemble variance) and the denominator is the mean squared error (i.e., equation 4.2 squared). N represents the number of ensemble members. As in equation 4.2 and 4.3, the spread computation in eqn. 4.5 is averaged over M grid points where the true state reflectivity is greater than 20 dBZ.

The prior analyses were compared to both the posterior and truth analyses to analyze the spatial structures of the prior analysis errors and the data assimilation increments. While posterior-prior (PO-PR) increments are useful to understand how the assimilation of observations impacts the state, the prior-truth (PR-TR) increments provide a sense of the accuracy of the state estimate throughout the model domain. Both sets of increments were averaged in time and space (along the convective line) over select periods within the OSSE to observe the systematic behavior of each set of increments. The line-averages are computed on a domain 200 km across that is translated each output time with the leading edge of the storms cold pool.

3.3 State-space RMSE

3.3.1 Truth simulation

The 3-hour truth simulation encompasses the development of 5 initially isolated cells into a quasi-linear convective system. The homogeneous environment used to initialize the model contains a vertical wind profile that produces splitting cells after approximately 30 minutes (Fig. 3.2a). By 90 minutes, the eight pairs of left and right cell splits merge to produce 4 cells, while the left (right) split of the northernmost (southernmost) cells remain isolated (Fig. 3.2b). These 6 cells proceed to merge into a linear convective system by the end of the simulation as their cold pools merge, forcing convection along the upshear edge (Fig 3.2c). The initially isolated cells are

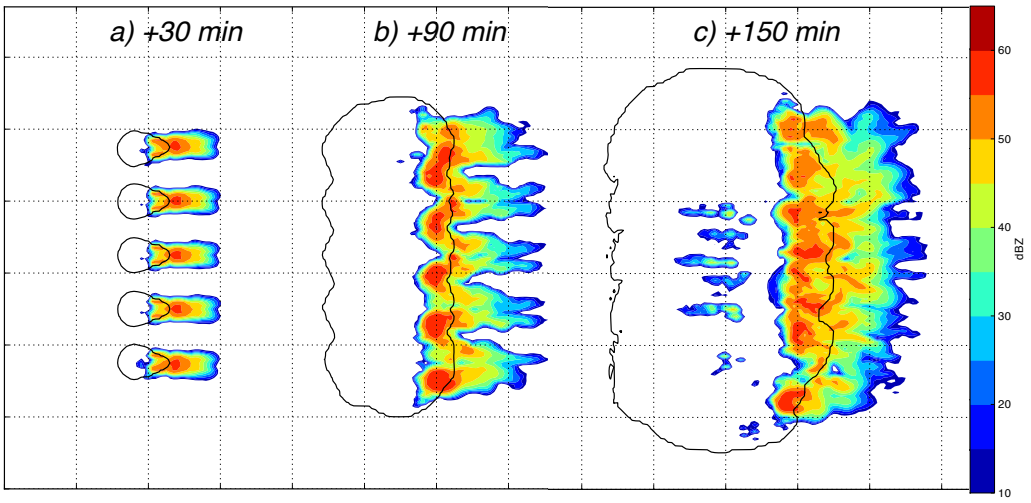

Figure 3.2: Column maximum reflectivity (color) and 0C perturbation theta (contour) from the truth simulation at a) 30 minutes, b) 90 minutes, and c) 150 minutes after initialization.

located approximately 50-80 km west of the radar at 30 minutes. By 90 minutes, the line of cells is co-located with the longitude of the radar, and by the end of the simulation, the squall line is 30-60 km east of the radar.

The process simulated by the model is representative of upscale growth observed in nature, when isolated multicells or supercells in close proximity merge into a convective line after several hours (this is a distinctly different process compared to convection forced along linear boundaries). By 3 hours, a well-developed cold pool is established with a vertically upright line of convection along the leading edge of the cold pool. While not used in this study, extending the simulation further (1-2 hours) results in a mature bow echo system.

3.3.2 LOC6V6 experiment

The LOC6V6 experiment is conducted using $r = 6$ km, a common choice in previous convective-scale data assimilation studies (e.g., Dowell et al. 2004; Dong et al. 2011). The horizontal grid spacing of the observations is 3 km, matching the model grid spacing. Observations extracted from the truth simulation at T+30 min are assimilated at the initial time in LOC6V6. No temperature bubbles are used in

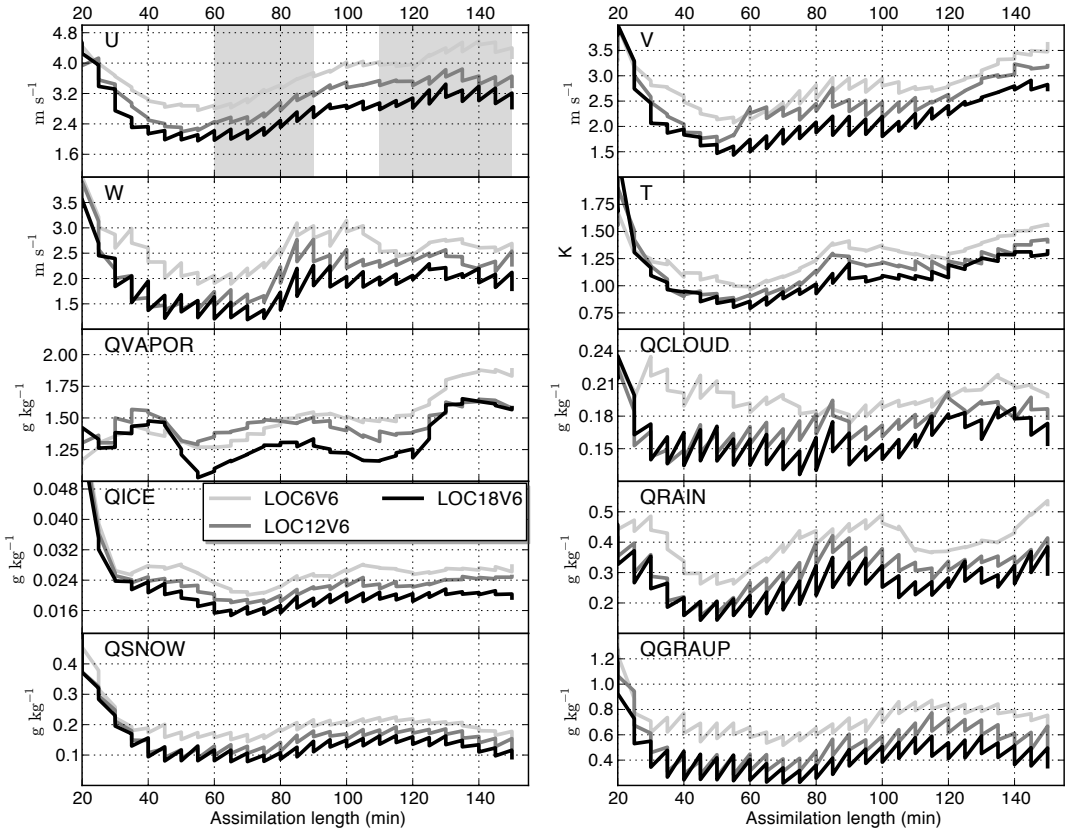

Figure 3.3: Domain-averaged state-space RMSE for LOC6V6, LOC12V6, LOC18V6 for the ten state fields updated by the data assimilation. Average includes grid points where the truth simulation reflectivity >20 dBZ. The spin-up period between 0-20 minutes is not shown. The two periods of cell mergers mentioned in the text are shaded for the U RMSE panel only.

the assimilation experiments to promote convective circulations. Thus, the model state takes approximately 20-30 minutes for the increments to stabilize as convection initially develops. Statistics are shown from this point forward (Fig. 3.3). Both the prior and posterior RMSE are plotted, hence the appearance of the sawtooth pattern. The RMSE decreases gradually for all state fields through the first 60 minutes of assimilation. This decrease is purely due to the assimilation of simulated radar data; the RMSE is not reduced due to the choice of boundary conditions, since open boundary conditions were used, or the choice of initial conditions, which is void of convection.

After 60 minutes, the RMSE increases until approximately 90 minutes; this 30-minute period coincides with the first set of cell mergers (shaded in Fig. 3.3). Afterward, the RMSE again stabilizes, although an increase in RMSE is evident in some state fields after 110 minutes (especially V, QVAPOR, and QRAIN). The increase in RMSE between 110 and 150 minutes (shaded in Fig. 3.3) is likely related to the second set of mergers that occur as the cells and associated cold pools merge near the end of the assimilation period. As the storm passes the radar location between 60-120 minutes, portions of the storm become unobserved due to the cone of silence above the radar. The effect of this on the RMSE statistics (and other analyses) is minimal since the unobserved portions of the storm are small compared to the size of the convective system. In addition, the increase in RMSE between 60 and 90 minutes occurs when the main convective cores are well-observed and only a segment of the anvil is unobserved. Thus, the increase in RMSE is thus most likely due to the cell merger process and not gaps in observations.

Comparing the velocity fields, the V field has the smallest RMSE compared to U and W throughout the assimilation period. This may simply be a reflection of a smaller overall V component of the horizontal wind compared to the U component, since the system is dominated by momentum in the west-east direction. While the horizontal wind is directly related to radial velocity, several of the other state fields (e.g. T and QVAPOR) are only changed through cross-correlations and the evolution of the dynamics during the model advance. For T and QVAPOR, most of the RMSE change occurs during the model integration. Among the other microphysical fields, QGRAUP has the largest RMSE, likely due to larger overall mixing ratios compared to QRAIN and QSNOW. In general, QGRAUP has a larger degree of variability in these experiments, which is evident in the slightly larger RMSE values.

3.3.3 Larger horizontal localization experiments

LOC12V6 and LOC18V6 both produce analyses with less error than LOC6V6 during the entire assimilation period (Fig. 3.3). During the first 60 minutes, the LOC12V6 and LOC18V6 state stabilizes to lower RMSE values more quickly than LOC6V6, but the LOC12V6 or LOC18V6 analyses are generally of similar quality during this period. Between 60-90 minutes, an increase in RMSE is also evident in the LOC12V6 and LOC18V6 analyses. Yet during this period, the RMSE of the LOC12V6 analyses increases more rapidly, thus LOC18V6 produces more accurate analyses. For most of the state fields, this result persists throughout the remainder of the assimilation period, although for some (e.g. QCLOUD, T, QRAIN), LOC12V6 and LOC18V6 produce comparable results near the end of the period. These results suggest that the underlying convective mode contributes to the ability of the EnKF system to produce accurate analyses. Specifically, in these experiments, a larger horizontal localization appears to be especially beneficial during the first period of cell mergers. This is also true for the vertical localization, as will be discussed in a following section.

The differences in r produce differences in posterior ensemble spread in the three experiments. As the localization cutoff is increased, state points are impacted by more observations, decreasing the posterior ensemble spread. The average prior state-space consistency ratio was computed to gauge the suitability of the spread within each experiment, similar to previous work (e.g., Snyder and Zhang 2003). This consistency ratio will be $\frac{N}{N+1}$, where N is the ensemble size, if the ensemble variance is equivalent to the error of the ensemble mean in these experiments (Murphy 1988). This optimal ratio is approximately 0.98 for the 50-member ensemble used herein. The consistency ratio within the storm (at Z >20 dBZ) increases during the first 50-minutes for each experiment and state field (Fig. 3.4) as the RMSE decreases. Afterward, the consistency ratio gradually decreases or remains steady through the end of the assimilation period as the RMSE remains steady or increases.

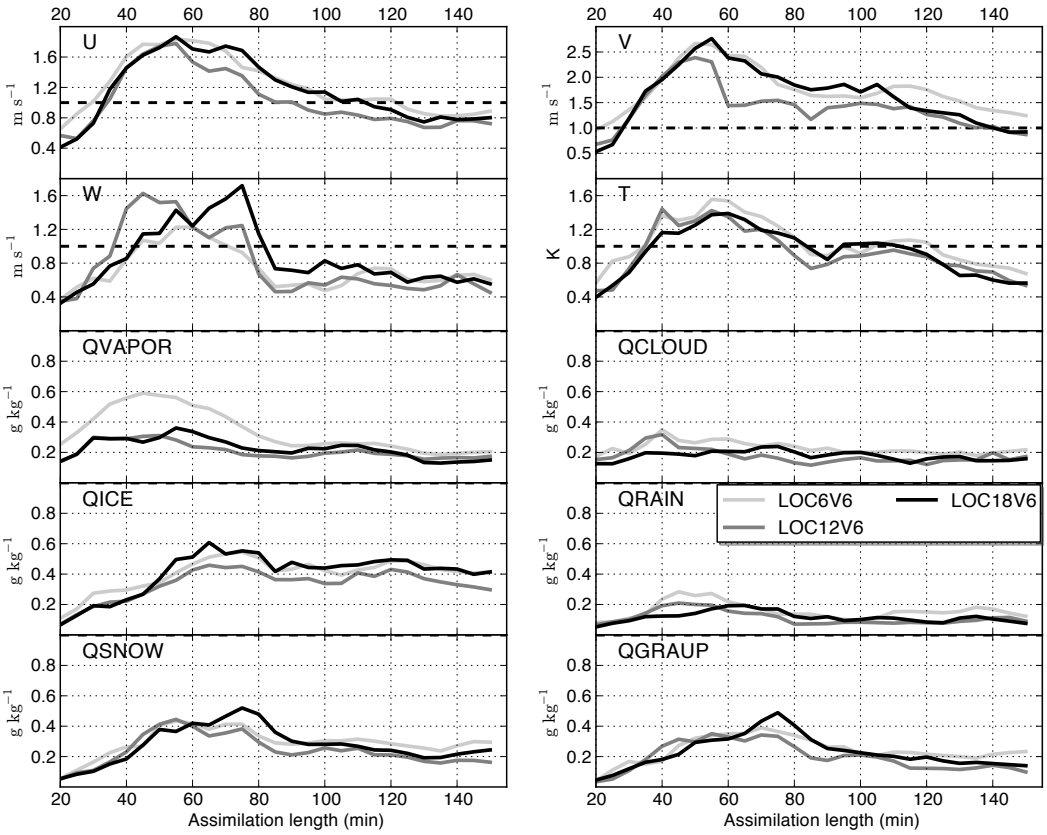

Figure 3.4: As in Fig. 3.3, but for prior analysis domain-average state-space consistency ratio. The dashed horizontal line indicates the optimal consistency ratio of 0.98.

As expected, the state fields altered by additive noise before the model advance (U, V, T, QVAPOR) have larger consistency ratios than the other state fields. For example, the U, V, and T consistency ratios are greater than the optimal ratio for the first half of the assimilation period in all three experiments, with values closer to the optimal ratio during the second half. The spread is likely too large for U, V, and T between 20 and 80 minutes, although this may be desirable to give more weight to the observations as the covariances develop in the ensemble. It is clear that efforts to implement adaptive spread adjustment techniques (e.g., Anderson 2009) in convective-scale EnKF studies will be necessary in the future.

3.3.4 Prior analysis vertical profiles

Vertical profiles of time-averaged RMSE and bias (equation 1 and 2) are shown in Figure 3.5. The biggest differences between the three experiments are found in the velocity fields. U, V, and W LOC18V6 RMSE are approximately 1 m s^{-1} smaller than LOC6V6 RMSE between 2 km and 12 km. The differences are similar between the experiments for U, but for V the differences increase with height. For all three experiments, the largest U, V, and W errors occur between 3 and 7 km, coincident with the strongest vertical velocities (not shown). Only small differences exist between the three T RMSE profiles. LOC18V6 has consistently smaller RMSE for QRAIN, QSNOW, and QGRAUP throughout the profile. The behavior of the biases between the 3 experiments isn't consistent among the state fields. The existence of some of these prior analysis biases, particularly the large negative bias that exists in the surface cold pool for T and QVAPOR will be discussed further in Section 3.4.

3.3.5 Sensitivity to vertical localization

The LOC12V6 vertical localization cutoff length, r_v, was changed in two additional experiments to assess the analysis sensitivity to changes in only the vertical

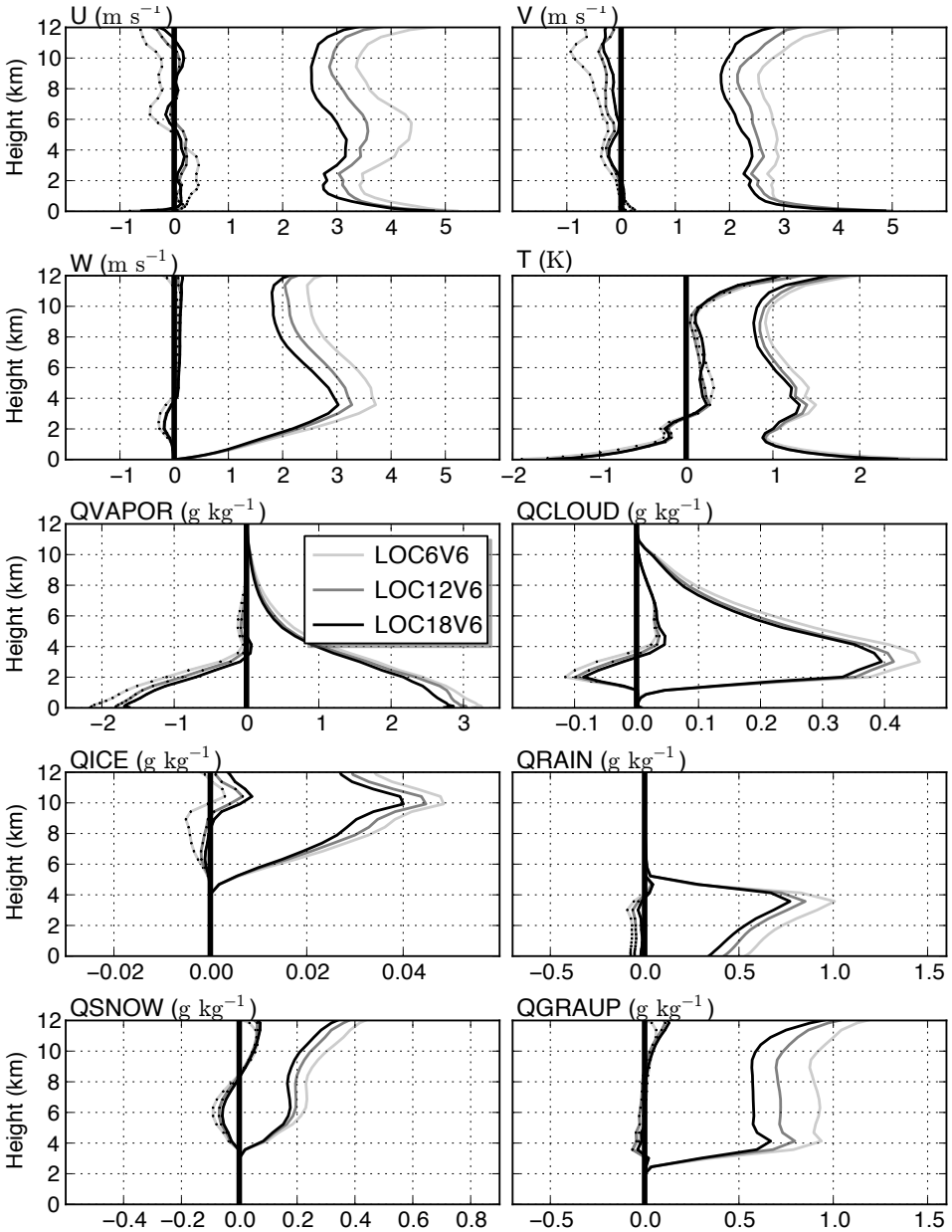

Figure 3.5: Average state-space prior analysis RMSE (solid) and bias (dotted) profiles for LOC6V6, LOC12V6, LOC18V6 for 10 state fields. Computed at prior state points where the true state reflectivity >20 dBZ from T+20 min through T+150 min.

localization. In these experiments, r_v was changed to 3 km and 12 km, while r_h remained at 12 km (LOC12V3 and LOC12V12). LOC12V6 from the previous section serves as a control experiment in this set of experiments.

During the first 60 minutes of assimilation, smaller values of r_v have either no appreciable impact or are slightly detrimental to the analyses (Fig. 3.6). The larger vertical localization appears to improve the spinup time for some of the fields during the first 40 minutes (e.g. W, QCLOUD). By 40 minutes, the differences in RMSE between the three experiments are minimal. Between 60-100 minutes, a noticeable difference in the RMSE between the three experiments emerges, with LOC12V3 having the smallest RMSE for most state fields (Fig. 3.6). As described earlier, an initial set of cell mergers occurs during this time period. The smaller vertical localization length scale in LOC12V3 appears to improve the state and covariance estimate during the time of cell mergers and interactions, and this benefit continues through the end of the assimilation period in most of the state fields.

In LOC12V3, many of the fields have assimilation increments that improve the state estimate. For example, assimilation decreases both the T and V analysis RMSE during the final 6 assimilation cycles (mean increments of approximately -0.1K and -0.2 m s^{-1}, respectively). In LOC12V6 and LOC12V12, the T and V analysis RMSE increases or remains unchanged after assimilation during these 6 cycles (mean increments of approximately +0.2K and +0.3 m s^{-1}, respectively). This is evident in other fields (e.g., QCLOUD, QVAPOR, QRAIN) as well, particularly during the second half of the assimilation period. Overall, the LOC12V3 experiment produces the smallest mean analysis RMSE.

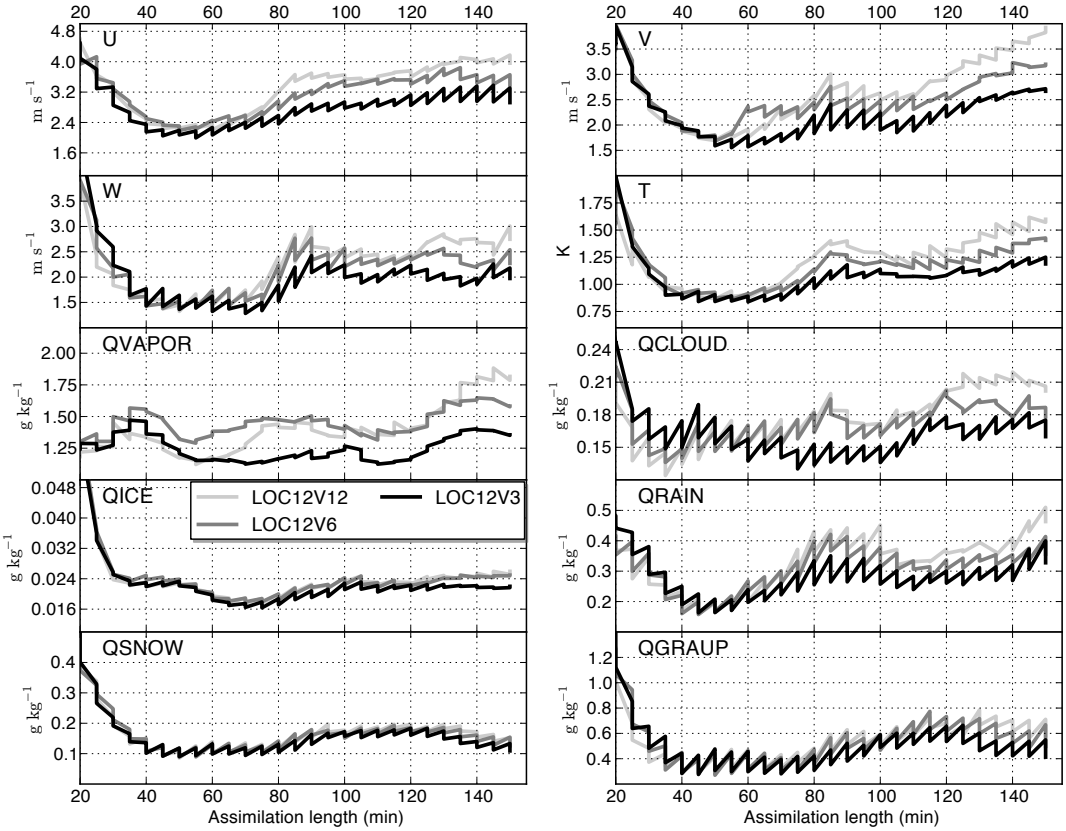

Figure 3.6: Same as Fig. 3.3, but for LOC12V3, LOC12V6, and LOC12V12.

3.4 Line-averaged, time-averaged increments

3.4.1 W and T increment patterns

Line-averaged, time-averaged (90-150 minutes) posterior mean prior mean (PO-PR) and prior mean-truth (PR-TR) updraft increments can elucidate systematic adjustments to the prior mean model state that are occurring within the convective system through the assimilation period. Line averages were taken in the y-direction from the northernmost to southernmost point along the convective line (from grid point 20 through grid point 110).

A negative-positive increment pair straddles the prior mean updraft (W) location in all of the experiments (Fig. 3.7). These adjustments indicate the location of the maximum updraft is, on average, displaced ahead of the truth updraft maximum. This is also evident in the PO-PR increments in the temperature field (Fig. 3.8). The negative updraft increments are collocated with negative temperature increments. At the surface below the location of the PO-PR updraft increments, positive PO-PR temperature increments exist. The assimilation, through the covariances between the radar observations and temperature field, is attempting to adjust the location of the cold pool by warming the leading edge by 0.5-1 K. These W and T PO-PR increments indicate the surface cold pool location is displaced eastward of the cold pool placement in the truth simulation, resulting in a low and mid-level updraft that is also biased eastward.

Compared to the truth simulation, not only is the assimilated convective system displaced eastward compared to the truth run, but the prior mean state surface cold pool is 1-3 K colder in all assimilation experiments. If the simulated cold pool propagates as a density current, then a colder cold pool will produce a simulated convective system with a faster forward speed (Wakimoto 1982) and will lead to the convective system in the assimilation experiments being slightly ahead of

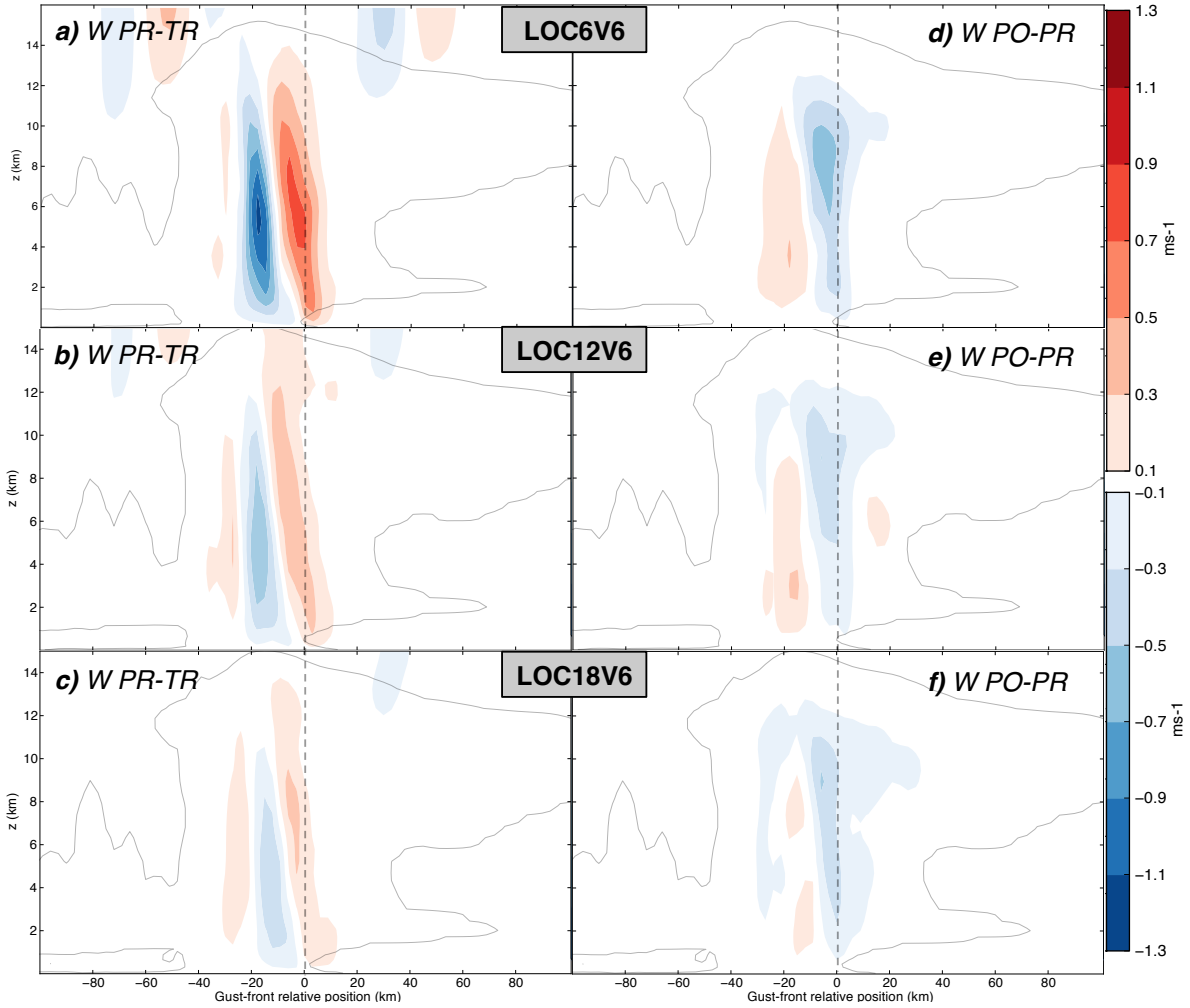

Figure 3.7: Line-time-averaged prior-truth (a-c) and posterior-prior (d-f) vertical velocity increments (shaded) from LOC6V6, LOC12V6, LOC18V6. The line averages are taken between grid point 20 and 110 in the y-direction and the time averages include analyses between 90 and 150 minutes. Also shown is the region where the combined cloud water and ice is greater than (gray contour). The average position of the leading edge of the surface cold pool is also shown (gray dashed line).

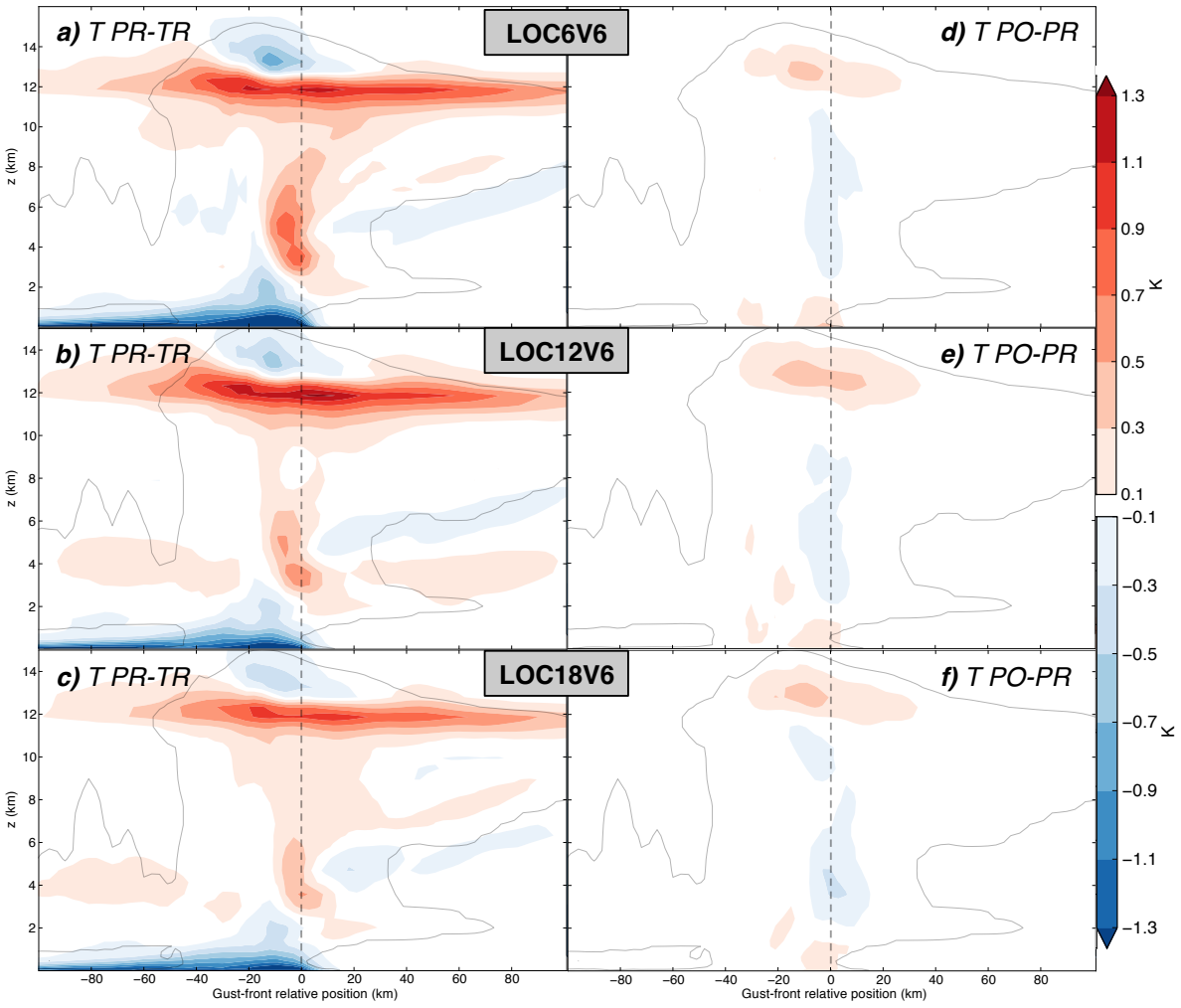

Figure 3.8: As in Fig. 3.7, but for temperature increments.

the system in the truth simulation. This relationship suggests that inaccuracies in cold pool temperature are producing the displacement error in the assimilated convective systems. The assimilation acts to correct this displacement error by adjusting the placement of the updraft and cold pool edge. Unfortunately, due to the lack of observations below 1 km (<10% of the total radar observations are located below 1 km), any adjustments to the cold pool must be accomplished exclusively through covariances between observations above the cold pool and state locations within the cold pool. Results suggest that the assimilation is unable to adequately adjust these fields, as the cold pool temperature and position errors remain throughout the assimilation period. Although, as will be discussed in the following section, changes in localization lessen the impact of this bias. While more observations below 1 km would improve the perfect-model results herein, in real cases with parameterized boundary layers, observations may still be unable to produce accurate analyses due to model error.

In addition to the errors in the mid-level updraft region, positive temperature errors exist in a band near the model tropopause (approximately 12 km) and negative temperature errors exist above the tropopause in the overshooting top. The assimilation attempts to correct the latter errors by adjusting the temperature upward in the mean posterior state (i.e. positive PO-PR T increments exist). Data assimilation is unable to correct the former temperature errors as evidenced by negligible PO-PR T increments. In both locations, observations are extremely sparse (e.g. <40 radar observations exist between above 12 km over the entire domain at each assimilation time), so the assimilation increments in the overshooting top region are likely dominated by cross-correlations that extend between observations in the updraft to the temperature field above the updraft. These correlations are unable to correct the errors near the tropopause, thus the errors persist and grow through the assimilation period. Errors in the horizontal wind field also exist at this level

(not shown). The wind errors supports the notion that these errors are primarily due to observational limitations, since radial velocity observations, if sufficiently dense, should be able to correct the wind errors in the upper-level outflow.

3.4.2 Differences between localization experiments

The patterns noted above are present in varying degrees in all of the assimilation experiments (Fig. 3.7, 3.8). In general, the spatial extent and magnitude of the PO-PR and PR-TR increments are the smallest in LOC18V6. For the T field, the PR-TR increments within the surface cold pool are smallest in LOC18V6. This is likely a contributing factor to the smaller W increments in the same experiment. In all three experiments, covariances develop between the assimilated radar observations and the state temperature field within the cold pool. The cold pool is a region that is inadequately sampled, so it is promising that the covariances are partly able to make appropriate corrections to the temperature errors near the surface, especially along the leading edge of the cold pool. The PO-PR increments are somewhat larger near the surface in LOC6V6, yet the magnitude of the positive increments in this region is similar between the experiments (since the vertical localization is not changing between these three experiments, the regions of the storm that rely on vertical covariances for corrections should be similar). The T errors near the tropopause are smallest in LOC18V6, although the PO-PR increments along the tropopause are similar between the three experiments.

3.4.3 Cold pool development sensitivity to additive noise

The convective system speed bias that develops in all of the experiments is a result of the prior state cold pool being colder than the truth simulation. In real-data experiments, cold pool biases often are attributed to errors in the model microphysics scheme (e.g., Dowell et al. 2011). In a perfect-model OSSE, the same microphysics

scheme is used in the truth and assimilation experiment, thus it is hypothesized that the errors are due to some component of the data assimilation process (e.g., spurious covariances between reflectivity and surface temperature) or additive noise. Dowell et al. (2011) demonstrated substantial sensitivity to the surface cold pool when the magnitude of additive noise was varied (compare their figure 9b and 9c). Specifically, their experiment which used additive noise values similar to those used herein produced a substantially larger cold pool, although, they were unable to verify which experiment was closer to reality. To assess the sensitivity of the present cold pool results to the additive noise, two additional simulations were conducted where the standard deviation of noise was reduced (compared to LOC12V12) to 0.25 (m s^{-1} and K; experiment LOC12V12N0.25) and 0.5 (m s^{-1} and K; experiment LOC12V12N0.5) beginning 20-minutes into the assimilation period. Retaining the larger values of noise during the initial 4 assimilation cycles is beneficial for storm spin-up (Dowell and Wicker 2009).

By the end of the assimilation period, PR-TR increments in LOC12V12N0.25 and LOC12V12N0.5 are smaller within the cold pool (Fig. 3.9a) and also reduce the PR-TR increments near the tropopause (not shown). Although the overall increments were smaller, PR-TR temperature increments of 1-2 K remained along the leading edge of the cold pool gust front in both experiments (Fig, 3.9a-b). At several locations, the prior state gust front in LOC12V12N0.25 was placed 10-15 km to the east of the gust front in the truth simulation, while other segments of the prior state gust front closely matched the truth. This explains the placement of the PR-TR increments near and ahead of the gust front compared to LOC12V12 in Fig. 3.9a. This also occurred in LOC12V12N0.5, although these were generally restricted to the northern half of the cold pool gust front; the southern half of the prior mean gust front matched well with the truth simulation in LOC12V12N0.5. The slightly better positioning of the gust front produced slightly smaller W PO-PR increments in LOC12V12N0.5

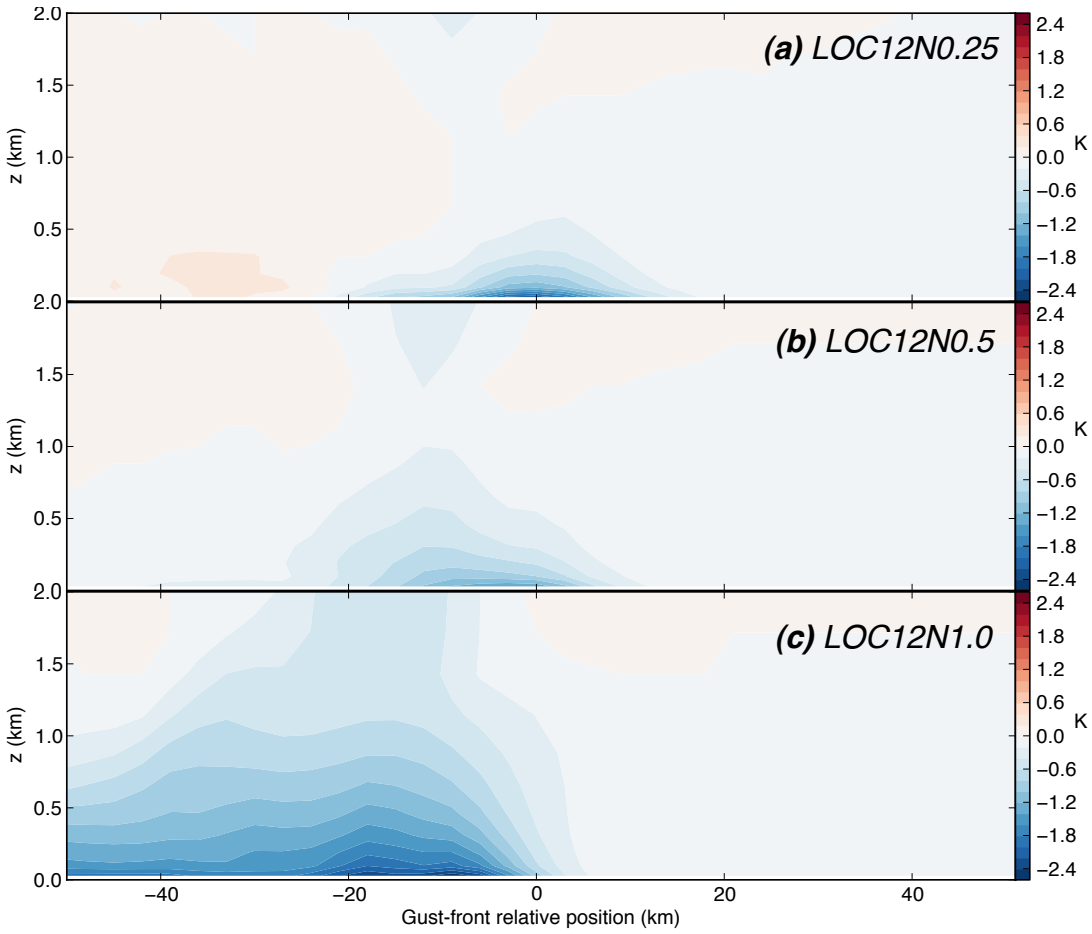

Figure 3.9: As in Fig. 3.8, but for PR-TR temperature increments below 2 km for (a) LOCH12V12N0.25, (b) LOCH12V12N0.5, and (c) LOCH12V12N1.0.

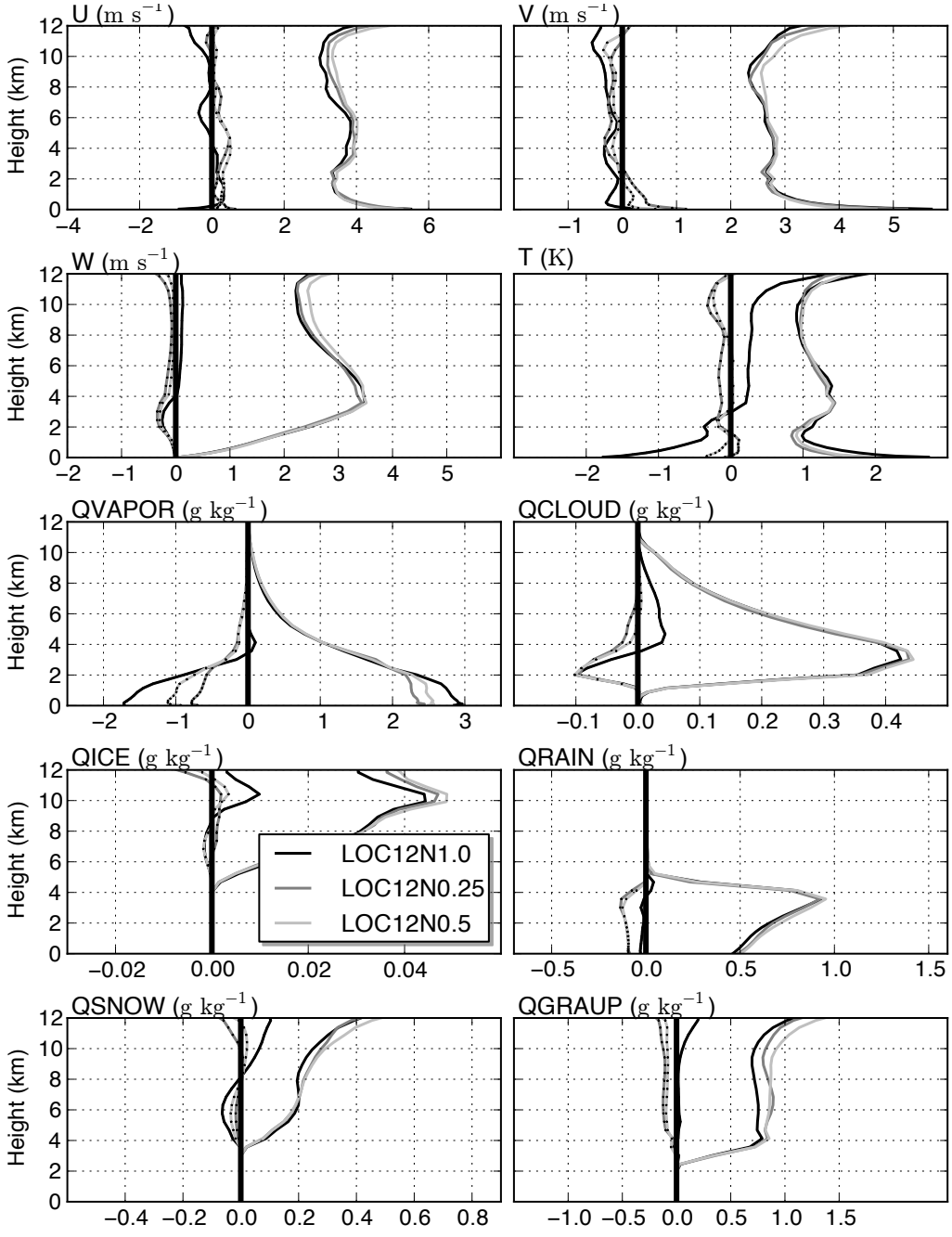

Figure 3.10: As in Fig. 3.5, but for LOCH12V12N1.0, LOC12V12N0.5, and LOC12V12N0.25.

compared to LOC12V12, yet the PO-PR increment pair remains in the time-line-averaged cross-sections (not shown).

These results suggest that using smaller values of noise produce improvements in the T state estimates, especially within the cold pool and anvil regions of thunderstorms. On the other hand, other state fields are slighly negatively impacted by smaller values of additive noise. For example, the average prior RMSE/bias profiles reveal that the microphysical fields have smaller errors using the original values for the standard deviation of the noise distribution (Fig. 3.10). This is also true for U, V, W, and T above the cold pool. Within the cold pool, U, V, T and QVAPOR RMSE are all reduced in LOC12V12N0.25 and LOC12V12N0.5 compared to LOC12V12, in addition to decreases in the large surface T bias. Further comment on the impact of additive noise on these results is provided in the discussion section.

3.5 Observation-state correlations

The results in the previous sections can be further interpreted by examining the covariance structures present within the OSSEs. Several previous studies have analyzed the covariance structures within convective storms using various techniques (e.g., Michel et al. 2011), primarily to develop covariance models for 3D-VAR applications. Herein, we apply a brute-force method by manually computing correlations and analyzing the spatial scales over which substantial correlations exist. Correlations are computed between the assimilated observations (radial velocity and reflectivity within the storm) and prior state fields (at the state locations where reflectivity is greater than 20 dBZ). For the present analyses, correlations are computed between radial velocity and the U, V, and W state fields, as well as reflectivity and the QRAIN, QGRAUP, and QSNOW state fields. These observation-state pairs were chosen due to their direct relationship to each other, although meaningful correlations exist between other observation-state pairs (e.g., reflectivity

and temperature). The prior state ensemble correlations drive the data assimilation update, and should provide insight into the characteristic spatial scales of the correlation field between the radar observations and state variables.

Observation-state distances and correlations are used to compute reverse cumulative density functions (RCDFs) that show the number of pairs greater than a given distance. These give an aggregate sense of the percentage of observations that fall at various distances away from the state points (or, alternatively, the percentage of state points that fall at a given distance away from the observation location). RCDFs were produced for the observation-state pairs with correlations greater than 0.4, 0.6, and 0.8. These thresholds were chosen to represent the length scales for mid- and high-end correlations that are not significantly affected by sampling noise.

3.5.1 3-D correlation structure

Figure 3.11a-c shows the RCDFs from the LOC12V12 60-minute prior analysis for all 6 observation-state pairs. Correlations greater than 0.4 exist for all observation-state pairs at distances up to 20 km (Fig. 3.11a), while those greater than 0.6 and 0.8 are generally restricted closer to the observation location (Fig. 3.11b-c). In the latter case, the number decreases quickly as the observation-state distance increases. For the microphysical fields, roughly half of the observation-state pairs with correlations greater than 0.4 are within 6.5 km (Fig. 3.11a). For the dynamic fields, this occurs at distances between 8 and 10 km. Thus, the radial velocity correlations with the dynamic fields (U, V, W) extend to further distances than the reflectivity correlations with the microphysical fields. This is also true for the larger observation-state correlations, although the RV,U correlations tend to behave similarly to the microphysical observation-state pairs.

Some insight into the performance of the various localization experiments can potentially be gleaned from these results. In LOC6V6, observations are not

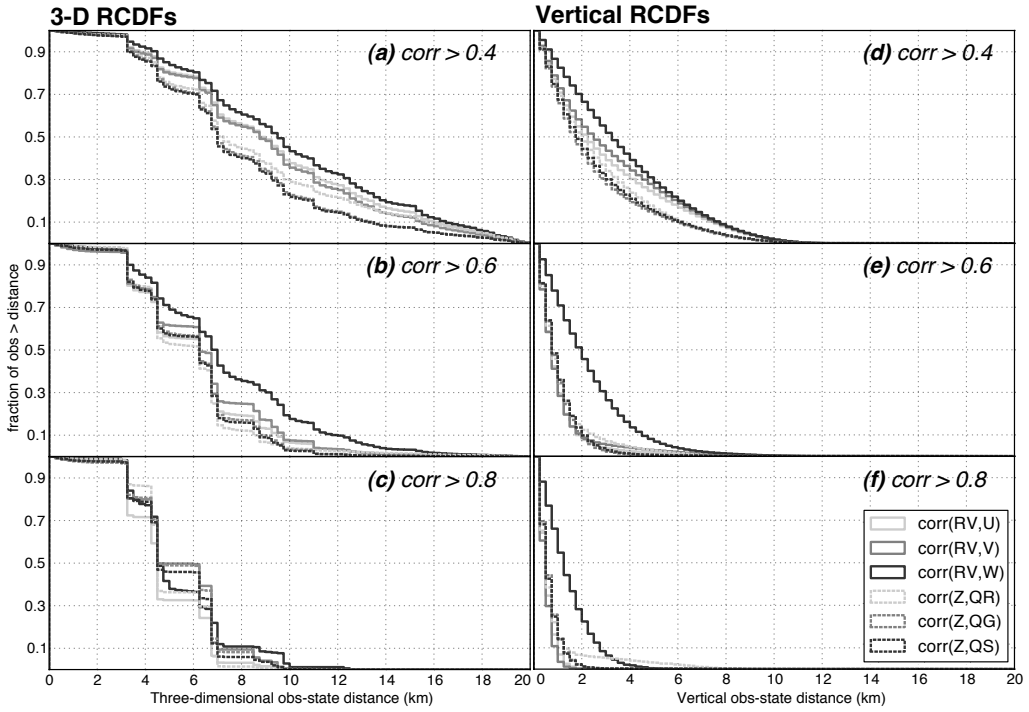

Figure 3.11: RCDFs for observation-state correlations greater than (a-d) 0.4, (b-e) 0.6, and (c-f) 0.8 using the (a-c) three-dimensional and (d-f) vertical observation-state distance from the 60-min prior analysis for LOC12V12.

permitted to update the state at distances beyond 6 km. Yet, only 20%-25% (40%) of the state points that are correlated at >0.4 with the radial velocity (reflectivity) observations lie within this radius (Fig. 3.11a; these results are consistent across LOC6V6 and LOC12V6). This may have proven to be detrimental to the LOC6V6 analyses, especially in data sparse regions. Increasing r_h to 12 km, as in LOC12V6, captures 70-75% of the state points that are correlated at >0.4 with the radial velocity observations (while capturing all of the state points with correlations >0.8). For reflectivity observations, roughly 90% (100%) of the state points with correlations greater than 0.4 (0.8) are within 12 km. While the localization choices in LOC18V6 permit the inclusion of an even greater fraction (nearly 100%) of the observation-state pairs, a larger fraction of noisier covariances less than 0.4 are also included (not shown). Thus, the improvement of the LOC12V6 analysis RMSE over LOC6V6

and LOC18V6 may be due to a closer match between the localization choice and the spatial scales present within the observation-state correlation field.

By the end of the assimilation period (150 minutes) the length scales of most observation-state pairs have increased compared to the correlations present at 90 minutes across all three experiments (not shown). This is hypothesized to be the result of upscale growth during the second half of the assimilation period. This will be examined further in section 3.5.3.

3.5.2 Horizontal vs. vertical correlation structure

To isolate the correlation spatial scales between the horizontal and vertical, the RCDFs in Fig. 3.11a-c were reconstructed based on the observation-state pair horizontal and vertical distance. The horizontal scales have similar structures to those shown in Fig. 3.11a-c, thus they are not shown. The vertical correlation scales are much smaller, especially at correlation thresholds of 0.6 and 0.8 (Fig. 3.11e-f). The length scales are similar for 5 of the 6 observation-state pairs, where approximately 90% (98%) of the pairs have vertical separation distances of less than 2 km for correlations greater than 0.6 (0.8). For the radial velocity-W pair, the length scales are longer at all correlation thresholds. Only 50% (70%) of these pairs have vertical separation distances less than 2 km for correlations greater than 0.6 (0.8). These findings suggest why LOC12V3 had the smallest RMSE between the three vertical localization experiments. A 3 km vertical cutoff contains a large majority of the observation-state pairs with correlations greater than 0.6, while reducing the number of observation-state pairs that are dominated by sampling noise.

3.5.3 Comparison with computed correlation length scales

Correlation length scales of the forecast errors from the OSSEs were computed using a technique described by Pereira and Berre (2006), hereafter PB06. While the RCDFs

provide some insight into the spatial structure of the observation-state correlations, the PB06 technique computes a local horizontal correlation length-scale for each model grid point for each state field. A domain median length-scale was computed for all model grid points where reflectivity was greater than 20 dBZ for each assimilation period beginning with the 30-minute prior ensemble analyses. The PB06 technique tends to produce large length-scales along the edges of the simulated convection. These length-scales were judged to be an artifact of the large horizontal derivatives of the state fields in these areas. The domain median, instead of average, length-scale was computed to reduce the overall influence of the length-scales at these grid points.

Some agreement exists between the RCDF and PB06 length-scales. First, the length scales associated with the microphysical fields from both methods are generally smaller than the dynamic fields (Fig. 3.12a) over the entire assimilation period. In addition, the changes in length scale during the assimilation period as the convection grows upscale (discussed in section 3.5.1) are also observed using the PB06 length scales. For instance, between 30 and 90 minutes, the PB06 length scales for U, V, and T increase by 300-500 meters. During this period, a more rapid increase occurs with QGRAUP and QSNOW between 30 and 60 minutes, followed by a period of little change for 60 minutes.

Differences are also evident between the two techniques. For example, the RCDFs produced using correlations larger than 0.4 indicate that the RV-W correlation length scale is larger than the RV-U and RV-V (Fig. 3.11a), while the W field PB06 length scale is smaller than the U and V length scales. In addition, separation is noted among the PB06 microphysical length scales while the RCDFs show no evidence of any difference between the three fields. Specifically, the QRAIN PB06 length scale is the smallest of the 7 fields for almost the entire assimilation period. This length scale evolves differently than the other 6 fields; instead of increasing or remaining steady during the 30-90 minute period, it gradually decreases during the entire 120-minute

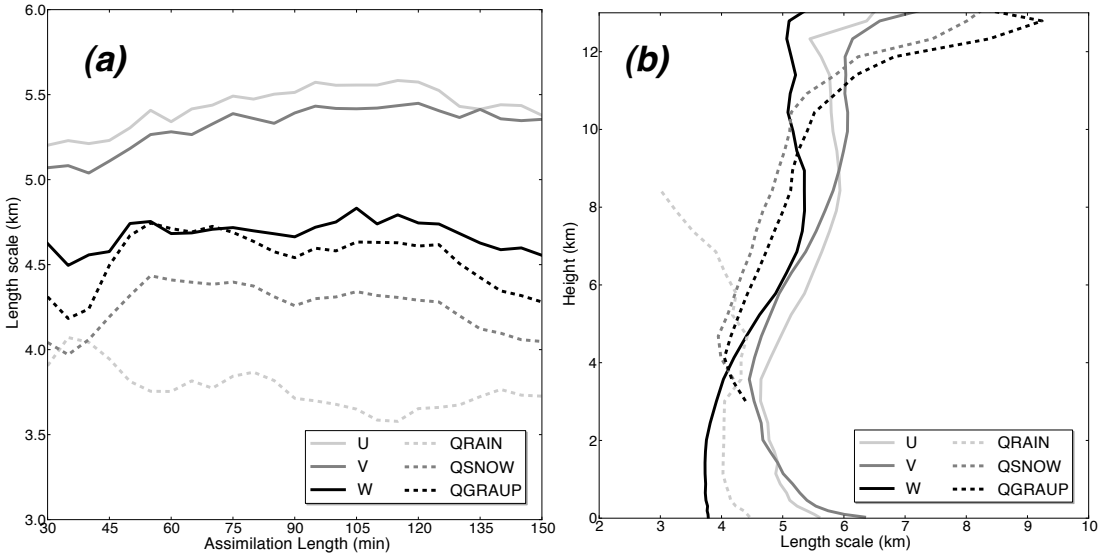

Figure 3.12: (a) Domain-averaged PB06 length scale during the assimilation period for 6 state fields from the prior analysis ensemble. Average includes grid points where the truth simulation reflectivity >20 dBZ. (b) Height-averaged PB06 length scale for 6 state fields from the final prior analysis ensemble.

period. During the final 30 minutes of the assimilation period, the PB06 length scales begin to decrease for all state fields except for QRAIN. There is no indication of this behavior using the RCDFs.

The median vertical profiles of PB06 length scale were created for the same 7 state fields for the final assimilation period (Fig. 3.12b). The U and V length scales are maximized near the surface and aloft (above 8 km) with a minimum near 3 km. The W length scale is smallest near the surface and is largest above 8 km. The QGRAUP and QSNOW length scales gradually increase from the mid-levels (they are not defined below this level) to the top of the storm. The length scale is smaller in the mid-levels than U, V, and W, but larger than these 3 fields above 11 km. The QRAIN length scale is fairly constant below 5 km and is 1-2 km smaller than the U and V length scales near the surface.

The PB06 length scales appear to have some substantial variability within the convective storms simulated in these experiments. For example, the W length scale at 6.3 km from the final prior analysis is plotted in Figure 3.13 (only points where

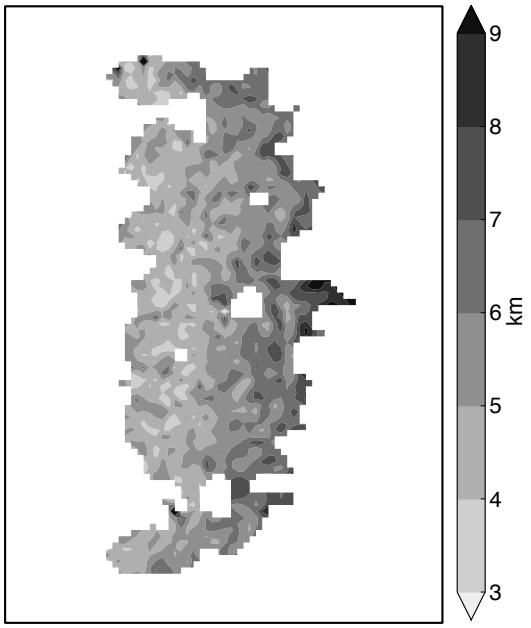

Figure 3.13: PB06 length scale from the final prior analysis for W at 7.4 km at points where the truth simulation reflectivity >20 dBZ.

the reflectivity is greater than 20 dBZ are shown). The length scale varies in the across-line direction by as much as 5 km, with larger length scales toward the front of the line and smaller length scales toward the rear. The PB06 length scales are more consistent in the along-line direction. While potentially revealing, further analysis of the spatial variability of the forecast error length scales is left to future work.

3.6 Ensemble forecasts

As another measure of the proper localization cutoff length, 60-minute ensemble forecasts are initialized from the final 50 model states at the end of the 150-minute assimilation interval. While analyses with smaller initial error should produce better forecasts, imbalance in the analysis also impacts the development of forecast error. Imbalance can be introduced due to suboptimal localization cutoffs, which could potentially hamper the ability to initialize forecasts with these analyses (Greybush et al. 2011). 60-minute ensemble forecasts were produced for LOC6V6, LOC12V6,

and LOC12V3 and LOC12V12. Ensemble mean state-space RMSE was used as a measure of the skill of the forecasts; the truth simulation was run for an addition 60 minutes to use as verification for the ensemble mean forecasts.

For most of the experiments and state fields, the rapid error growth occurs during the first 15-30 minutes (Fig. 3.14). For some fields, RMSE continues to increase through the end of the 60-minute forecast period (e.g., U, QRAIN) while for others the RMSE is steady (e.g., QVAPOR), or even decreases in the 30-60 minute period (e.g., W, QCLOUD). The model states become progressively more similar among the experiments during the forecast period. For example, the range of RMSE values from the ensemble mean U analysis at the initialization time is roughly 2 m s^{-1}, while the error differences in the 60-minute forecast of U are less than 1 m s^{-1}.

For most of the forecast period, the RMSE evolves similarly in time between the experiments, with the forecast error a function of the initial analysis error (i.e., the experiments with smaller 60-minute forecast RMSE have smaller initial RMSE). For example, LOC12V3 generally produces the most accurate analyses, and the 60-minute forecasts benefit from this more accurate initial state.

While the sensitivity of the forecasts to localization is unique to this study, previous studies (e.g. Aksoy et al. 2010; Dawson et al. 2012) that examined convective-scale ensemble forecasts using the EnKF have found similar results in terms of error growth, although their statistics were computed in observation-space. A significant difference in this work is the apparent decrease in error evident in several state fields as described above. This may be due to the simplistic view of convective-scale forecast quality used herein; other more advanced verification metrics (e.g. Schwartz et al. 2010) could provide a different view of these findings.

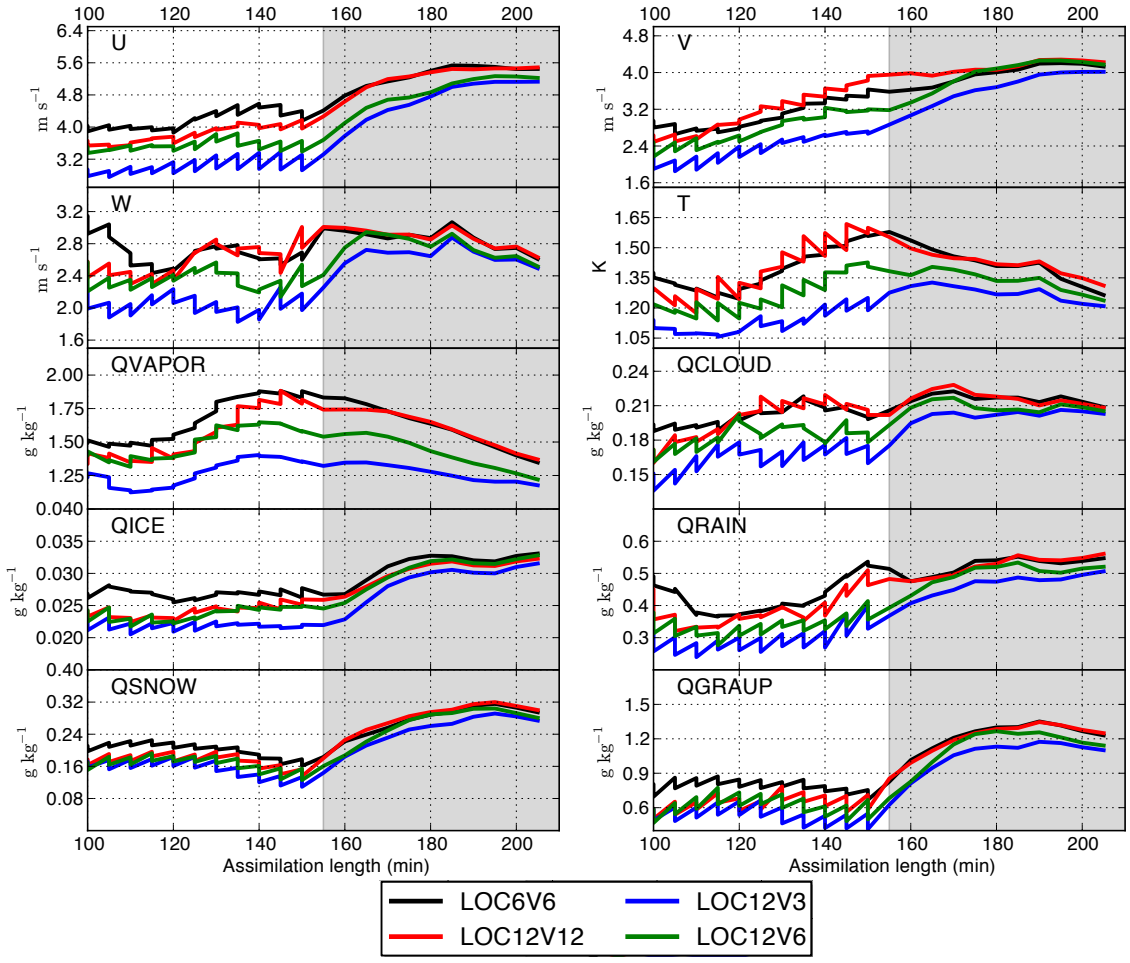

Figure 3.14: RMSE for ensemble mean forecasts initialized with the final posterior ensemble from LOC6V6, LOC12V12, LOC12V3 and LOC12V6 for 10 state fields. The last 12 analysis cycles are shown (unshaded), in addition to the 60-minute ensemble forecasts (shaded).

3.7 Summary and Discussion

Part I describes the results of several OSSEs to assess the impact of covariance localization for radar data on EnKF analyses of a developing convective system. Sensitivity to both horizontal and vertical localization was assessed via RMSE statistics, assimilation increments, and prior-truth increment structures. To interpret the results further, the prior correlations between the assimilated radar observations and state variables within the convective storm were computed and analyzed both in aggregate (RCDFs) and on a local scale (PB06 length scales). Further, ensemble forecasts were produced from the final EnKF analyses to provide another measure of the impact of different localization options on the analyses.

The results herein suggest that when assimilating radar data with the temporal and spatial resolution typically provided by a WSR-88D, a larger horizontal (r_h = 12-18 km), but smaller vertical (r_v = 3 km) localization cutoff (using a typical Gaspari-Cohn localization function) than previous studies could be beneficial, especially in cases where the convective evolution is dominated by thunderstorms interactions, mergers, and/or upscale growth into a linear convective system, as was the case in the present simulations. These findings run counter to the localization choices used in all previous convective-scale EnKF literature, especially the vertical localization cutoff, where the vertical length scales of the covariance structures for radar observations were seemingly limited. When considering the choice of localization, factors such as ensemble size and inflation should be considered. For example, a larger localization length scale would likely be beneficial when employing a larger ensemble (Houtekamer and Mitchell 1998).

Some discrepancies between the previous work and this present work exist that are worth mentioning. For example, Tong and Xue (2005) noted that r = 6 km produced the best results after decreasing the ensemble size from 100 to 40 members, which is similar to the ensemble size used herein. One difference between Tong and

Xue (2005) and this work is the model grid and observation density (2 km vs. 3 km). Another difference is the mode of the simulated convective system. Previous convective-scale EnKF studies employing both OSSEs and real-data experiments, primarily have focused on producing analyses of an individual thunderstorm cell (e.g., tornadic supercells). While these studies have laid the foundation for convective-scale EnKF and proved its viability, the scenarios depicted in these studies are only a subset of the large variety of convective modes observed in nature. Thus, more work is needed to demonstrate how those results generalize to other types of convective systems and forms of severe weather.

The results of the current experiments were obtained with the assumption of a perfect model. In reality, model error (e.g. due to microphysical parameterizations) plays a significant role in producing accurate analyses and will likely modulate the exact results shown herein. Specifically, model errors can produce erroneous covariance structures, which could change the results reported herein. In Part III, real-data experiments are performed to examine the sensitivity of the analyses to radar data localization in a setting where model error is not neglected.

The results herein reiterate the difficulties with analyzing the surface cold pool, even in perfect-model OSSEs. Producing accurate analyses of the cold pool structure near the surface is essential in order to produce accurate forecasts of the speed and structure of cold-pool driven convective storms whose behavior and severity are largely controlled by the thermodynamic properties of the surface cold pool, even more so than other convective modes (e.g., supercells). Yet, a paucity of observations near the surface, sensitivities to assimilation parameters such as additive noise, and microphysical scheme errors (in real-data experiments) will make this a challenging task.

Herein, covariance localization was achieved by applying an isotropic weighting function to all ensemble covariances, based on the observation distance from the

state location being updated. While this type of localization is easy to implement and has proven to be useful, the present results indicate a need for more flexible, adaptive localization options that vary based on location, state variable type, time, and dynamical considerations. For example, during a time when cell mergers and interactions were occurring the OSSE using the largest localization radius produced the most accurate analyses. In addition, the covariance length scales increase as the overall system grows upscale and the covariances exist substantial spatial variability. Furthermore, in the present experiments, radial velocity observations are related to the dynamic state variables of U, V, and W on slightly larger length scales than reflectivity observations are related to the microphysical variables. This result is consistent with several studies that have developed heterogeneous background error covariance models for 3D-VAR applications (e.g. Michel et al. 2011). While the present results may provide guidance for Gaspari-Cohn localization choices, they also underscore the need for techniques that determine the meaningful prior ensemble covariances dynamically during the data assimilation step (e.g., Bishop and Hodyss 2009a,b; Chen and Oliver 2009).

Chapter 4

Part II: Surface and radar data assimilation for the 29 May 2012 convective event

In this chapter, a real-data DA study of a convective event was performed as a complement to Part I to explore issues not readily studied in the OSSE framework. First, routine and mesonet surface observations are assimilated prior to CI, followed by both surface and WSR-88D data after CI. During the former period, assimilating mesonet data with the EnKF is anticipated to provide more accurate forecasts of CI, a considerable challenge in the operational forecasting community (Weckwerth and Parsons 2006; Kain et al. 2013). A more accurate representation of the boundary layer and mesoscale surface boundaries should extend benefits into the latter time period as well, leading to improved forecasts once radar DA begins. In the post-CI period, surface DA can constrain convectively-generated surface features, such as cold pools, that may be important in predicting later convection (e.g. Stensrud et al. 2009b).

As summarized in sections 2.4 and 2.5, using the EnKF to assimilate sub-hourly surface (i.e. 5-minute mesonet data) and radar observations with the goal of improving explicit predictions of CI has not been documented in the literature. Further, the 5-minute assimilation of surface and WSR-88D data with the EnKF following CI extends previous work that performed surface and radar DA with variational techniques (e.g. Schenkman et al. 2011a) and on mesoscale grids that are used as initial and boundary conditions for storm-scale experiments (e.g. Yussouf et al. 2013). The complexity of the convective event, complete with multiple storm modes, merging cells, and upscale growth, is also a novel aspect of the current work. It is hypothesized that surface DA, in addition to radar DA, provides additional impact

for these more complex convective events compared to events consisting of isolated convection (e.g. supercells). To gauge the success of these strategies, 3-hour ensemble forecasts are initialized at hourly frequencies throughout the DA period. Few studies (e.g. Snook et al. 2012) have assessed the efficacy of the EnKF DA techniques with multiple sets of ensemble forecasts initialized with EnKF analyses. A variety of open questions exist regarding the use of surface and radar observations for forecasts of both CI and post-CI convective evolution. Several overarching research questions that the real-data experiments are designed to address are provided below:

- Does sub-hourly assimilation of mesonet observations improve explicit forecasts of CI and later convective evolution?
- What is the relative impact of assimilating both surface and radar data on 0-3 hr forecasts of convection?
- What challenges exist in producing EnKF analyses of a complex convective event?
- What biases are present that may reduce the effectiveness of assimilating surface and radar observations?
- In light of the results from Part I, are the ensemble forecasts sensitive to the choice of radar data localization?

A summary of the event and the justification for its selection are provided in section 4.1 and 4.2, respectively. The experiment design is described in section 4.3, while a review of other convection-permitting forecasts for this event is provided in section 4.4. The effects of surface DA on ensemble forecasts of CI are analyzed in section 4.5, while section 4.6 discusses the impact of radar and surface DA on 0-3 hr ensemble forecasts of convective evolution.

4.1 Case summary

4.1.1 Synoptic and mesoscale environment

On the synoptic scale, a shortwave trough was located across the northern Great Plains and western Great Lakes region at 12 UTC 29 May 2012 (Fig. 4.1a). In the wake of this trough, shortwave ridging and height rises were occurring across the Rocky Mountains. An axis of stronger flow aloft (50-60 knots at 300 hPa) associated with the subtropical jet stream was oriented across the southwestern U.S. into the southern Great Plains. By 00 UTC 30 May 2012 (Fig. 4.1b), the shortwave trough had moved eastward and height rises were occurring across the central and southern Great Plains. Associated with this ridging and the subtropical jet stream, there was modestly strong west to northwest flow above the 700 hPa level across the southern Great Plains.

By the afternoon of 29 May 2012, a stationary front was draped across northern OK and southern KS, demarcating the boundary between cooler, dry air to the north associated with the high developing behind the northern U.S. trough, and more moist, warm air to the south across OK (Fig. 4.2). The 00 UTC 30 May 2012 radiosonde launched from Norman, OK was representative of the environment in the warm sector (Fig. 4.3). This sounding sampled the environment downstream of the developing convection to the west of the launch location, but was uncontaminated by precipitation or cloud layers. During the afternoon hours, an area of low pressure and associated dry line developed across western TX, primarily due to strong diabatic heating within the dry air mass across the High Plains of eastern NM and western TX (Fig. 4.2). By early evening, the air mass to the east of this dry line and south of the aforementioned stationary boundary was characterized by moderate to strong instability. Mixed-layer convective available potential energy (MLCAPE) values ranged from 2000-4000 J kg^{-1} across the warm sector, due to the presence

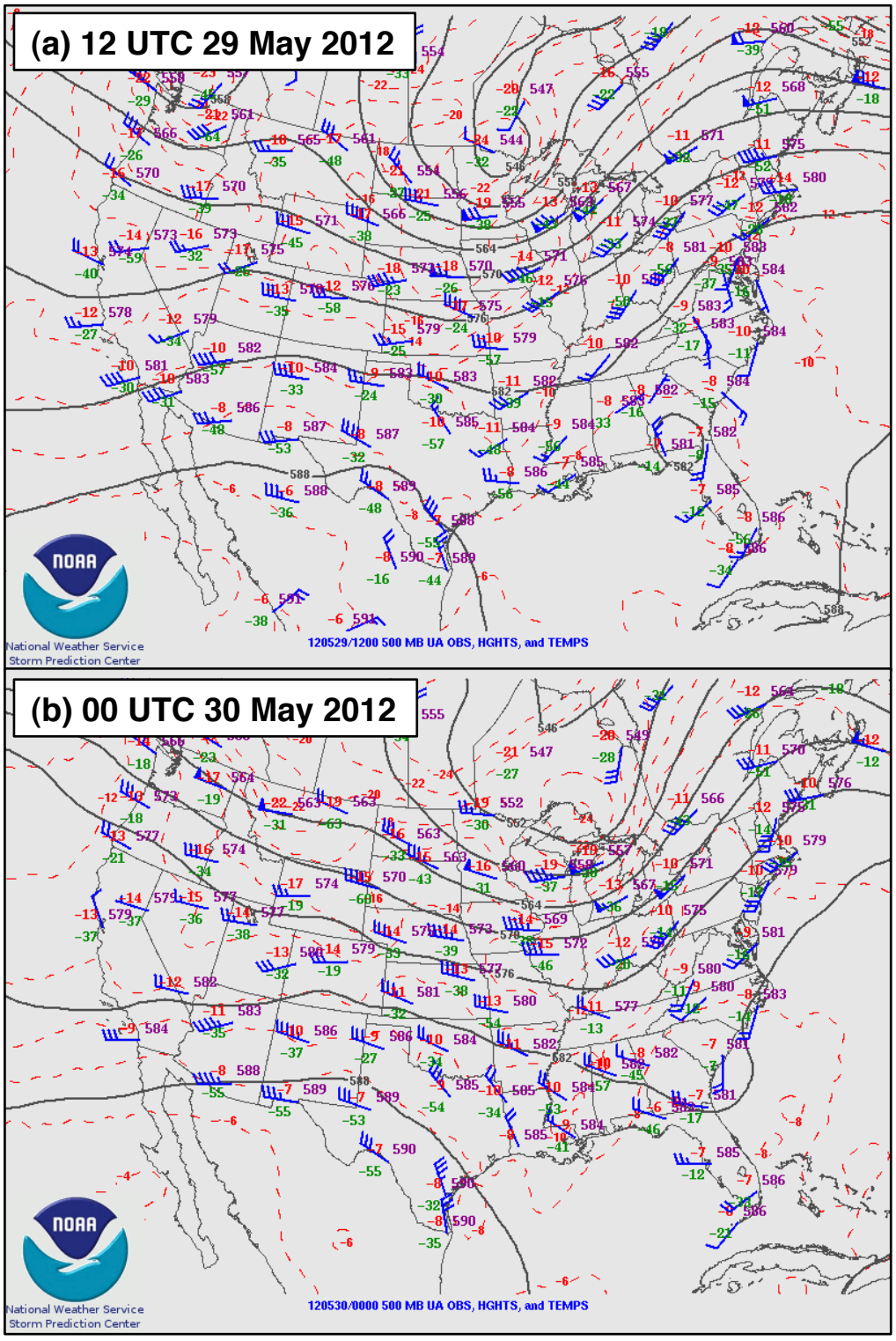

Figure 4.1: 500 hPa upper-air observations, heights, and temperature at (a) 12 UTC 29 May 2012 and (b) 00 UTC 30 May 2012.

of steep mid-level lapse rates (>7.5 C/km; Fig. 4.3) and a moist boundary layer. Deep-layer shear was supportive of organized convection (0-6 km shear >40 kts; Fig. 4.3) due to the presence of the 30 knots of northwest mid-level flow overlaid ontop of southeasterly surface flow.

Thus, although there was limited synoptic scale support, instability and wind shear were supportive of organized convection across most of OK by mid-to-late afternoon on 29 May 2012. The presence of 0-6 km shear at or above 40 knots and a long, straight hodograph was conducive for the development of both left-moving and right-moving supercellular convection (Klemp and Wilhemson 1978). Modest values of low-level shear (0-1 km shear was <10 knots; Fig 4.3) and lifted condensation level (LCL) heights (LCLs were near 1.5 km AGL) supported primarily non-tornadic supercells, with large hail and severe wind the most likely threats (Thompson et al. 2003).

4.1.2 Convective evolution

CI occurred between 21 UTC and 22 UTC, in three regions near the surface dry line and stationary boundaries: 1) near the intersection of the stationary front and dry line in southwest Kansas (Fig. 4.4a), 2) in northcentral Oklahoma near and to the north of a dry line bulge, and along a well-defined NW-SE oriented boundary layer roll (Fig. 4.4c), and 3) in southwestern Oklahoma along the dry line, in the presence of several boundary layer rolls (Fig. 4.4c). In addition, a failed CI attempt occurred in northern TX (Fig. 4.4a). For the remainder of this work, these four regions, and their associated CI events, will be denoted CI1-4.

Between 22 UTC and 00 UTC several supercells developed within CI1-3. In CI1, an intense supercell moved southeast into NW Oklahoma, while weaker convection developed in southwest KS (Fig. 4.5b). This supercell weakened after 23 UTC and dissipated shortly after 00 UTC (Fig. 4.5d). The two areas of CI

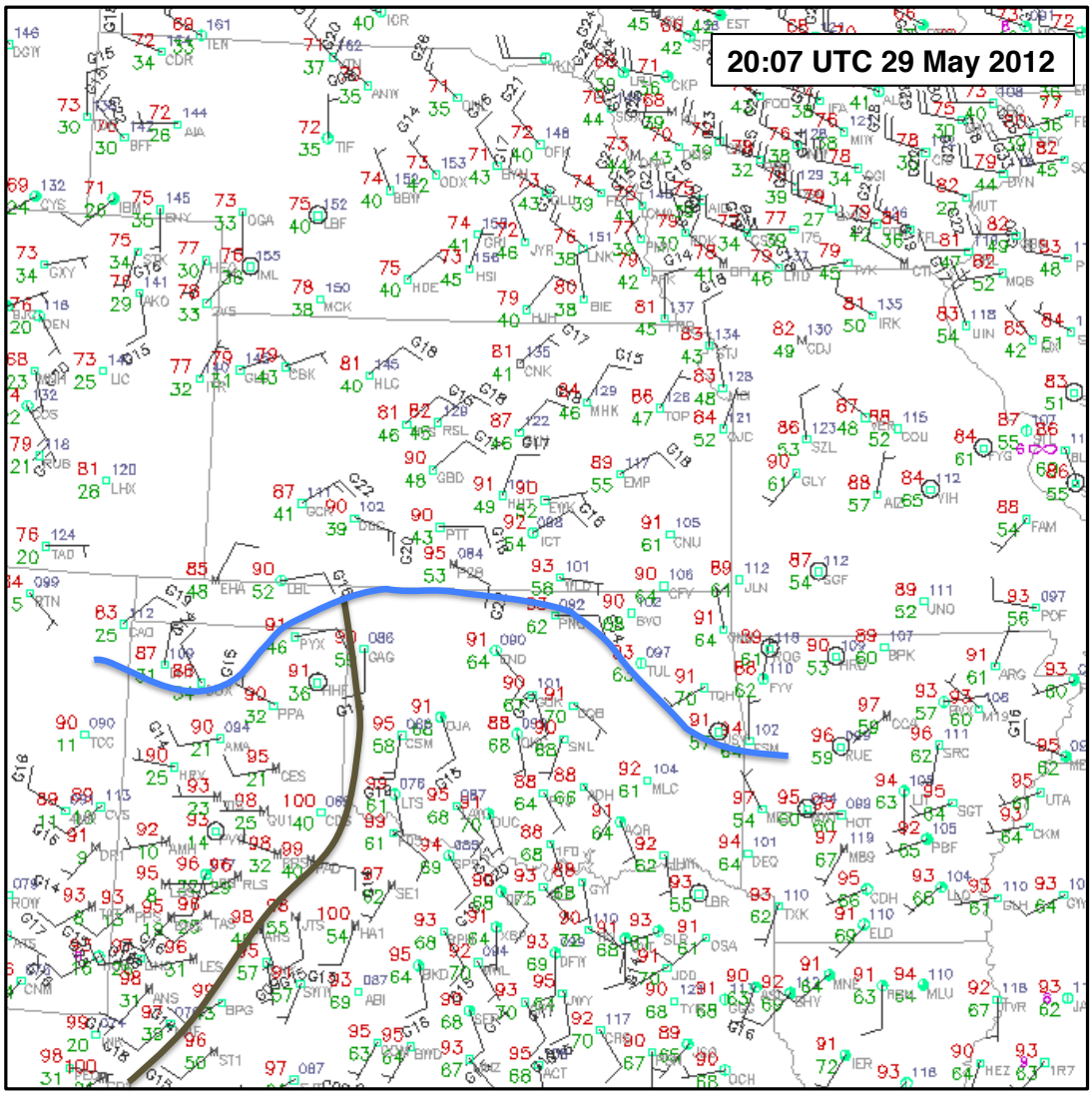

Figure 4.2: Surface map valid at 20:07 UTC 29 May 2012. Solid lines indicate estimated positions of stationary boundary (blue) and dry line (brown). Observations are plotted at each station according to station model conventions.

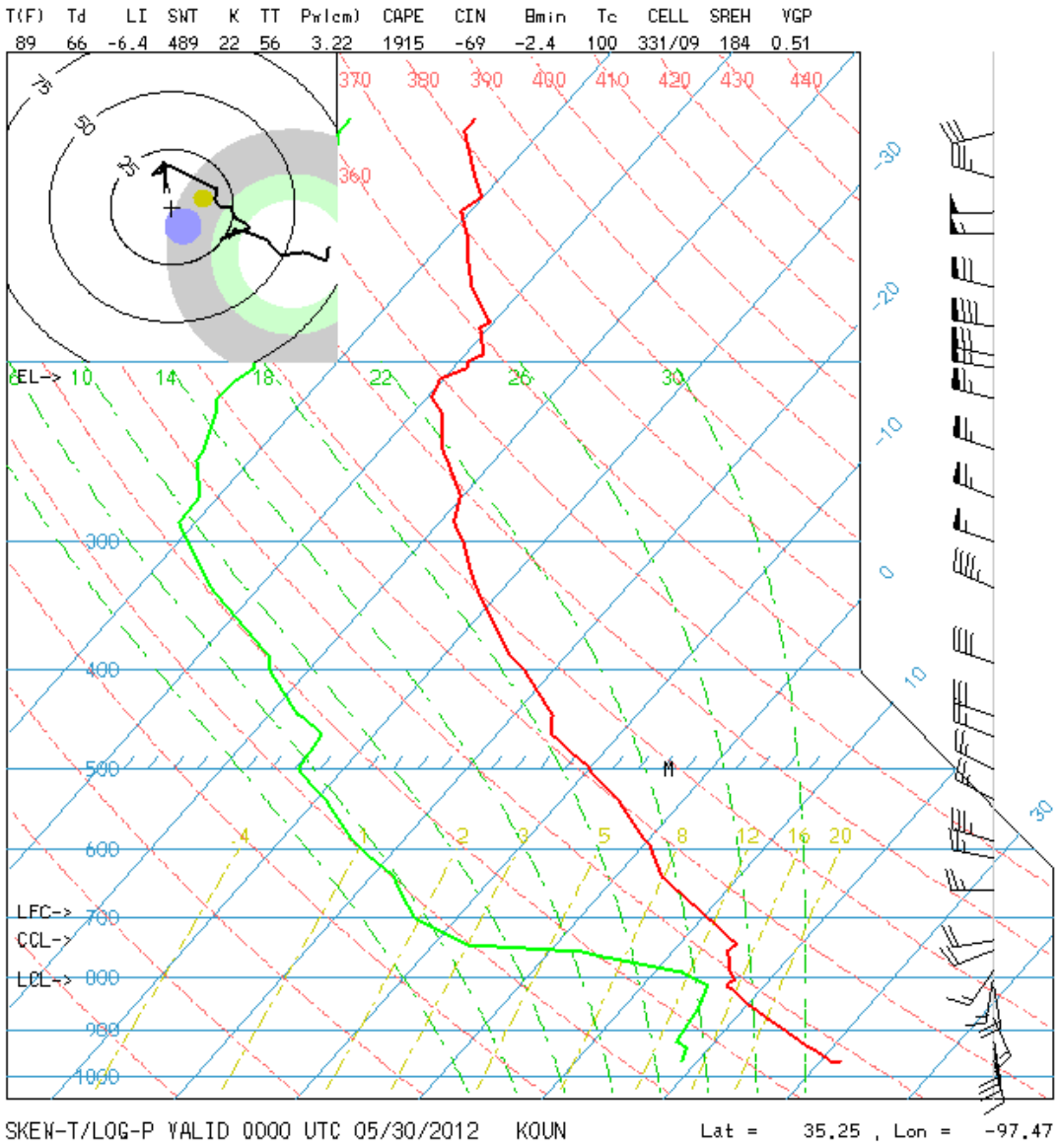

Figure 4.3: Skew-T graphic showing the 00 UTC 30 May 2012 sounding from KOUN.

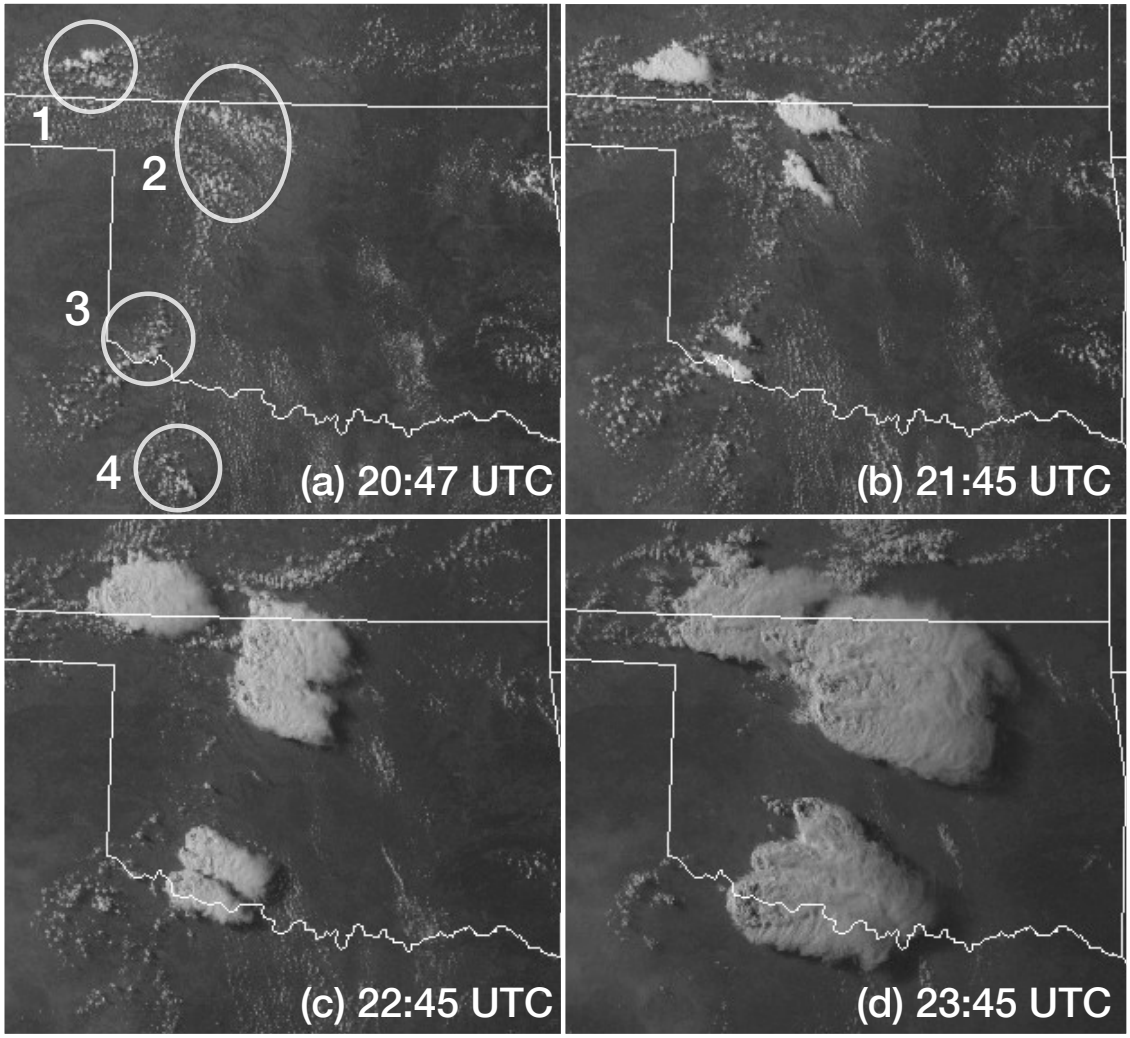

Figure 4.4: Visible satellite image from (a) 20:47 UTC, (b) 21:45 UTC, (c) 22:45 UTC, and (d) 23:45 UTC. Circles 1-3 denote areas of CI, while circle 4 denotes an area of CI failure, within the domain. These regions are referred to as CI1–4 in the text.

within CI2 (Fig. 4.5a) developed into three primary convective cells by 23 UTC (Fig. 4.5b). By 00 UTC, many convective cells had developed within CI2, readily forming on outflow boundaries from early convection (Fig. 4.5c). Of note in this region are several intense supercells on the southern end of CI2 that moved toward the south, passing within and to the west of the Oklahoma City (OKC) metropolitan area (Fig. 4.5d). Within CI3, two isolated supercells developed and matured between 22 UTC and 00 UTC (Fig. 4.5a-c). The northern storm produced a left split that persisted as it moved northeastward (Fig. 4.5d) toward OKC. The southern right-moving supercell moved toward the southeast, crossing the Red River into northern Texas (Fig. 4.5d). The left-split continued to move northeast until merging with the southeastward moving convection from CI2. This merger occurred within the OKC metropolitan area shortly after 01 UTC.

After 01 UTC, a large convective system developed across central and eastern Oklahoma, consisting of convection from CI2 and new convection developing to the east of CI2. This MCS was composed on the western end by several supercells that surged toward the south under the influence of an expanding surface cold pool. Although under the influence of an expanding cold pool, the convection within the western end of the MCS retained characteristics of discrete supercells for several hours, while the eastern end of the MCS across eastern Oklahoma was more linearly organized and developed after 00 UTC. The southern right-moving supercell from CI3 persisted until shortly after 03 UTC, moving southeast into the region south of Wichita Falls, TX.

4.1.3 Storm reports and damage

All forms of severe weather occurred across Oklahoma on 29 May 2012, including hail, wind, and a brief tornado (Fig. 4.6). Nearly $500 million in damage occurred within Oklahoma during the 29 May 2012 convective event (NCDC 2012). This was

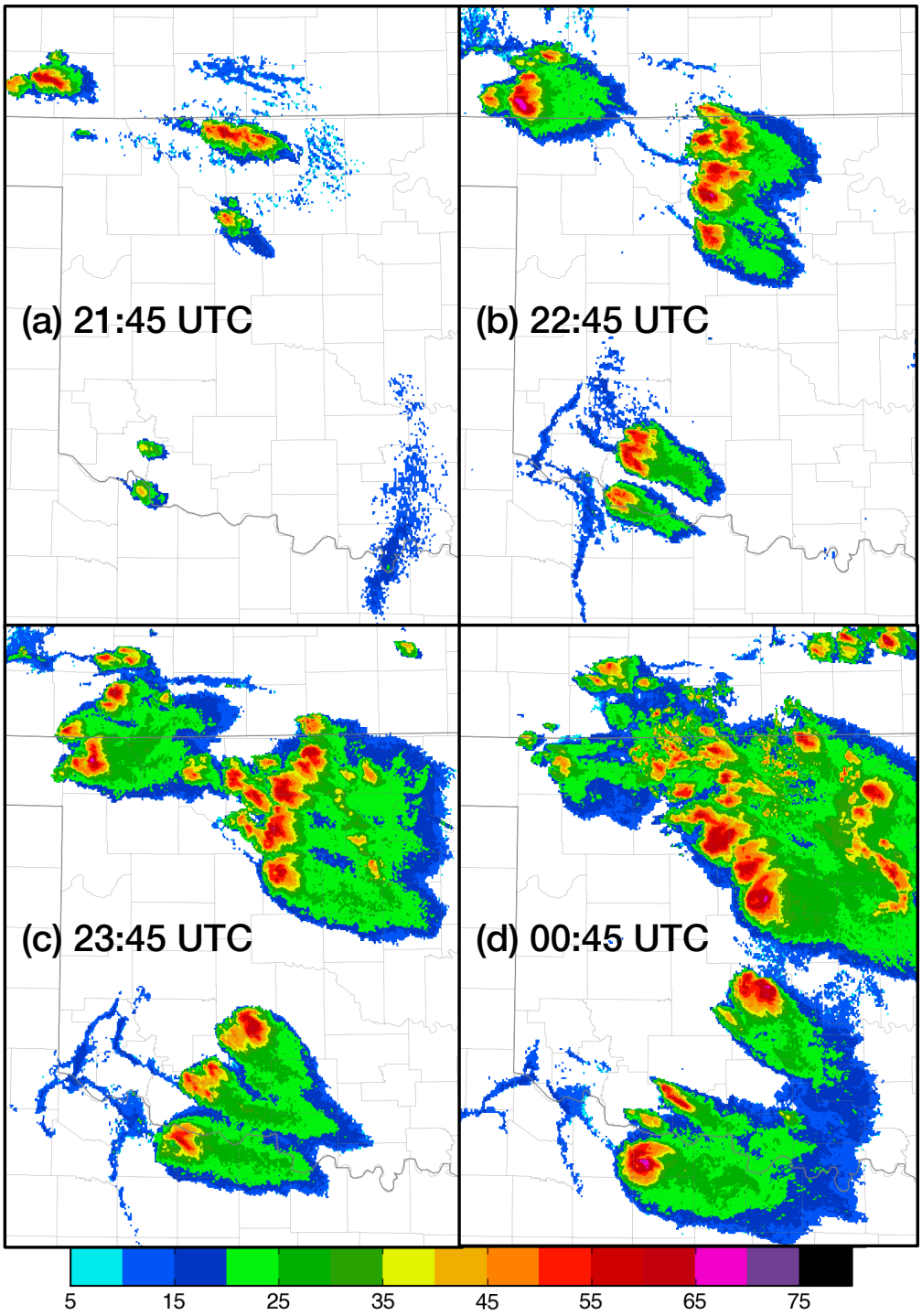

Figure 4.5: Composite reflectivity (dBZ) at (a) 21:45 UTC, (b) 22:45 UTC, (c) 23:45 UTC, and (d) 00:45. (a)-(c) correspond to the same time as Fig. 4.4b-d.

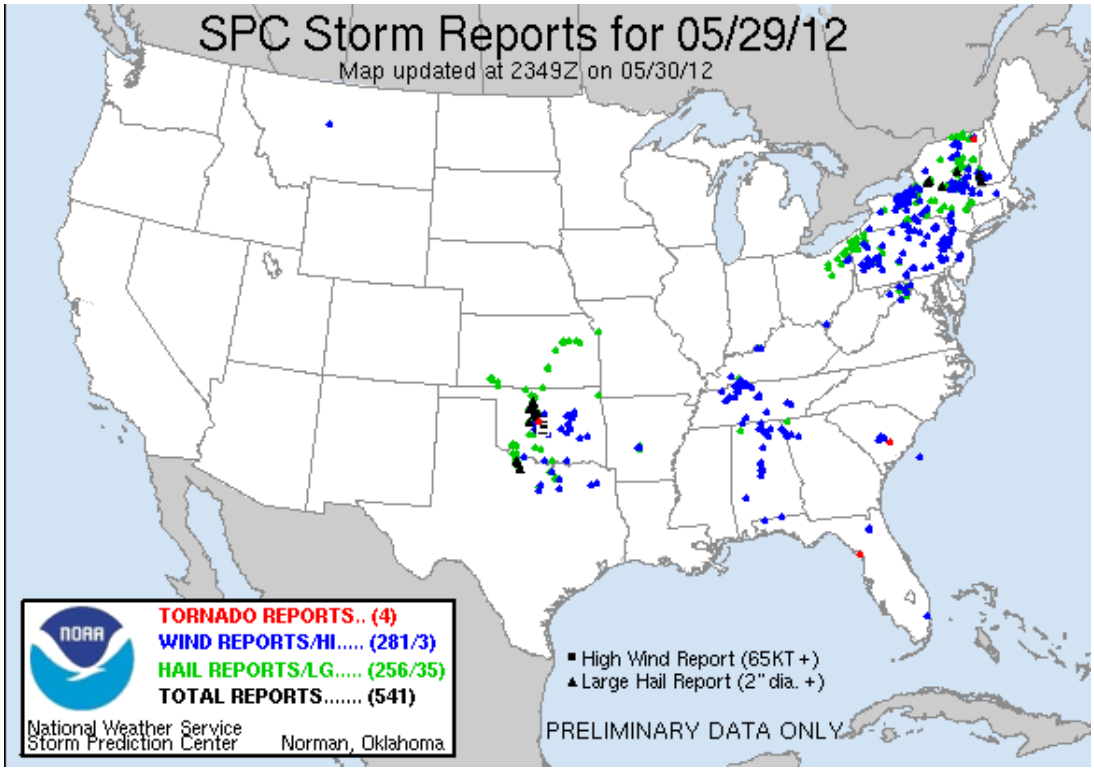

Figure 4.6: Wind (blue dots), hail (green dots), and tornado (red dots) reports from 12 UTC 29 May 2012 through 11:59 UTC 30 May 2012. Significant severe weather reports are denoted by black squares (wind gust >65 knots) and black triangles (hail >2" or more in diameter). The cluster of reports near OKC includes several significant severe weather reports.

primarily associated with 5" diameter hail that was observed with the CI2 supercells and the left-moving CI3 supercell within the OKC metropolitan area. In addition, a short-lived EF-1 tornado occurred west of OKC with one of the CI2 supercells. Wind gusts to 80 mph were also reported with the CI2 supercells, while 60-70 mph wind gusts occurred with the convective line that developed after 00 UTC across eastern Oklahoma and northeastern Texas.

4.2 Case selection

The 29 May 2012 convective event was chosen to test various hypotheses about the beneficial use of surface and radar data for predictions of complex convective events. The characteristics of this event described in the previous section were desirable for several reasons. First, synoptic scale forcing was relatively weak, thus initiation was driven primarily by forcing along surface boundaries. It is anticipated that surface data assimilation could potentially provide greater benefits for these types of cases. Second, the lifecycle of the convective system was constrained to a reasonably sized domain (1000 km x 1000 km), minimizing computational expense. Third, a high-quality mesoscale background was available for this case. Finally, a variety of storm modes were observed during the event, including left-moving and right-moving supercells, cell mergers, and upscale growth into an MCS. This variety provides a unique opportunity to test the ability of EnKF assimilation systems to provide accurate forecasts of more complex convective events.

4.3 Experiment Design

4.3.1 Model configuration and initial ensemble

The assimilation experiments used a model grid with a horizontal grid spacing of 3 km, while the vertical grid contains 40 vertical levels and is 16 km deep. The domain is 945 km (315 grid points) in the north-south direction, and 885 km (295 grid points) in the west-east direction, and was centered over Oklahoma (Fig. 4.7). This configuration keeps most of the convection within the domain and away from the lateral boundaries. The initial and boundary conditions for the 50-member, 3-km ensemble originated from downscaled 15-km mesoscale analyses taken from a continuously cycled 50-member EnKF analysis system run in real-time at the National Center for Atmospheric Research (NCAR) during the Spring of 2012. This system employs the Advanced Research WRF version 3.3.1 as the forward model (WRF; Skamarock et al. 2008) and the ensemble adjustment Kalman filter assimilation algorithm within the Data Assimilation Research Testbed (DART; Anderson et al. 2009) software package (described in Part I). The NCAR mesoscale ensemble was initialized on 30 April 2012 and observations were assimilated every 6 hours through the current period of interest. Further details about the design of the NCAR EnKF mesoscale analysis system are provided in Appendix A.

To create the initial conditions for the 3-km ensemble, the posterior ensemble state valid at 18 UTC on 29 May 2012 was downscaled onto the 3-km model grid. To produce boundary conditions, the 15-km mesoscale ensemble was advanced to 06 UTC on 30 May 2012 (i.e. 12 hour forecast), with boundary conditions from the 18 UTC GFS. The same GFS-based boundary conditions were used for each member. This should not lead to a detrimental reduction in spread in the 3-km ensemble due to the interior domains placement far away from the 15-km domain boundaries. The

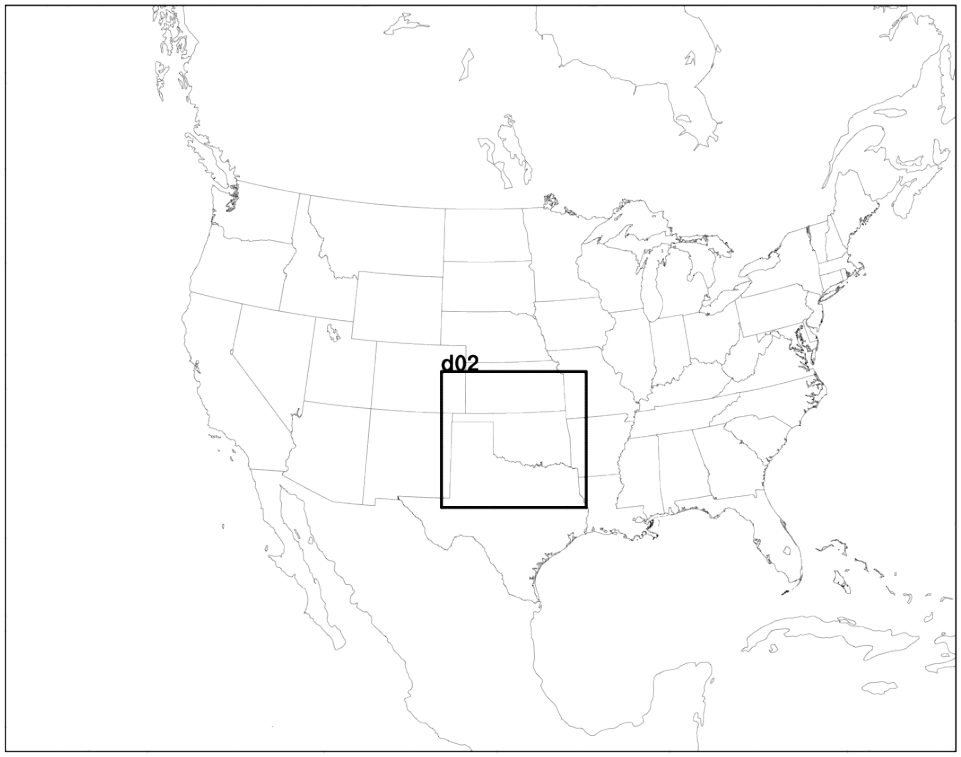

Figure 4.7: NCAR EnKF analysis domain (outer domain; 15 km horizontal grid spacing) and nested inner domain (3 km horizontal grid spacing).

program *ndown*, part of the WRF package, processed 30-minute output from the 15-km ensemble to produce initial conditions at 18 UTC and boundary conditions every 30-minutes. The initial ensemble for all experiments consists of the 50 downscaled ensemble states valid at 18 UTC on 29 May 2012. A control experiment with no data assimilation was performed beginning from this ensemble and extending to 06 UTC 30 May 2012 (i.e. a 12-hour free forecast). Other experiments were setup to assimilate various combinations of surface and radar data between 18 UTC and 00 UTC.

Version 3.3.1 of the WRF model was used for the assimilation and forecast experiments. The Morrison double-moment scheme (Morrison et al. 2009) was used to parameterize cloud microphysics. This scheme was adapted from WRF version 3.4.1, since the later version includes a computation for reflectivity (details are located in Appendix B). The Morrison scheme predicts mass mixing ratios for cloud water,

WRF setting	Value
Horizontal grid	315 x 295, Δx = 3 km
Vertical grid	40 levels, p_{top} = 50 hPa
Cumulus scheme	None
PBL scheme	MYJ
Microphysics scheme	Morrison
Radiation (LW)	RRTMG
Radiation (SW)	RRTMG
Land-surface scheme	NOAH

Table 4.1: WRF model settings

cloud ice, rain water, snow, and graupel, as well as the number concentration for rain, ice, snow, and graupel. Boundary layer parameterization is implemented with the Mellor-Yamada-Janjic (MYJ) scheme (Janjic 1994). The Rapid Radiative Transfer Model for Global climate models (RRTMG; Iacono et al. 2008) handled shortwave and longwave radiation. Other details of the model configuration are provided in Table 4.1. The WRF model was modified to output hourly-maximum values of fields useful for the diagnosis of convection in high-resolution model output (Kain et al. 2010). In addition, 2-m temperature, 2-m mixing ratio, 10-m wind speed, and composite reflectivity were saved at 5-minute intervals for verification with observations.

4.3.2 Observation sources and processing

Surface data were obtained from the NOAA Global Systems Division Meteorological Analysis and Data Ingest System (MADIS; Miller et al. 2005). These data included observations from standard METAR sites and various mesoscale networks (mesonet observations). The MADIS system does not archive Oklahoma Mesonet data (although it is available in real-time), thus these data were added separately.

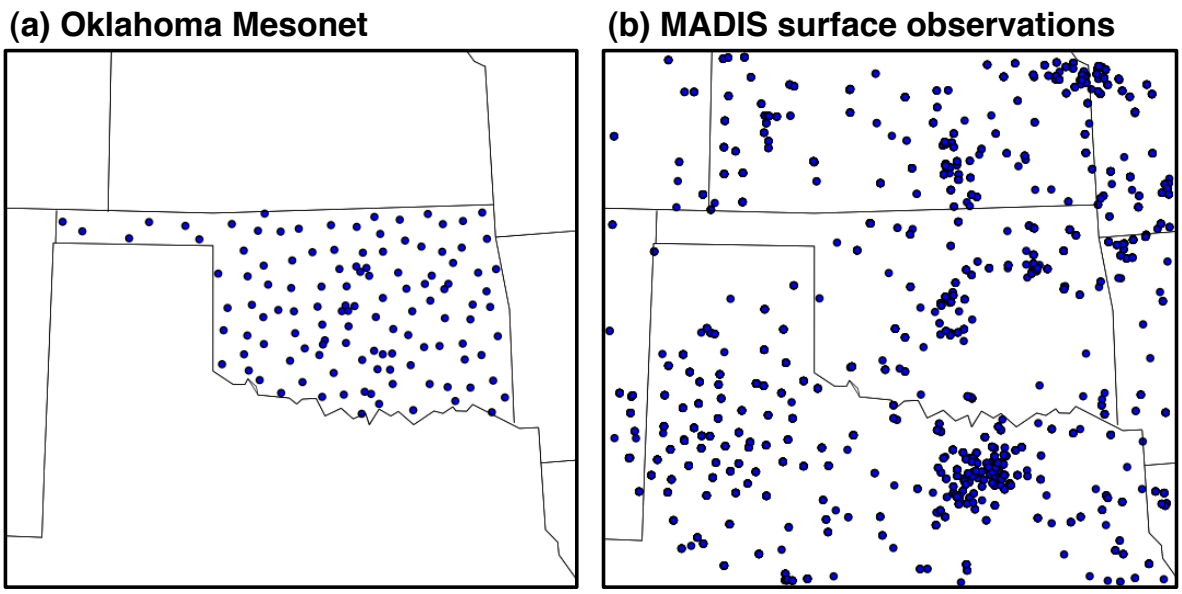

Figure 4.8: Distribution of surface observations from the (a) Oklahoma mesonet and (b) MADIS surface observation dataset at 18 UTC 29 May 2012 within the 3-km nested domain. (b) includes only those MADIS observations that passed all MADIS quality-control checks.

Temperature, dew point temperature, altimeter, and the horizontal wind components (U and V) were assimilated. A rigorous quality control procedure, including spatial and temporal consistency checks, was applied to the observations within the MADIS system. Only those MADIS observations that have passed all MADIS quality control checks were assimilated in the present experiments. Figure 4.8 shows the spatial distribution of surface observations at 18 UTC 29 May 2012. Observation error standard deviation for U, V, temperature, and altimeter observations are shown in Table 4.2. Dew point observation uncertainty is computed with the root-sum-of-squares method using temperature and relative humidity uncertainties (following Lin and Hubbard 2004, their eqn 4a). Observation errors for MADIS mesonet data were assigned to be slightly larger than METAR and Oklahoma mesonet data due to the variety of mesonet data sources with unknown error characteristics within the MADIS dataset.

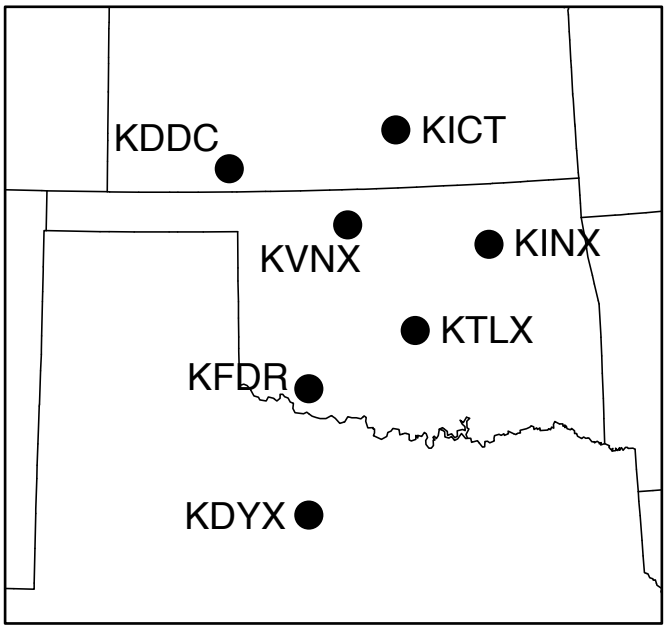

Figure 4.9: Locations and identifiers of the 7 WSR-88Ds used in the radar DA experiments.

Data source	**U** $(m\,s^{-1})$	**V** $(m\,s^{-1})$	**T** (K)	**Altimeter** (hPa)
METAR	1.75	1.75	1.75	0.75
OKMESO	1.75	1.75	1.75	0.75
MADIS MESO	2.5	2.5	2.5	1.5

Table 4.2: Summary of observation errors for surface datasets. Dew point temperature observation error assigned using Lin and Hubbard (2004).

In addition to surface data, reflectivity and radial Doppler velocity observations from 7 WSR-88Ds (KICT, KDDC, KVNX, KINX, KTLX, KFDR, KDYX; Fig. 4.9) were assimilated into several experiments. The radar observations were objectively analyzed using a single-pass Barnes (1964) objective analysis scheme onto a regularly spaced 6 km grid using the Observation Processing and Wind Synthesis (OPAWS) software. As in previous studies (e.g. Dowell et al. 2004; Xue et al. 2006), the observations were kept on the two-dimensional conical sweep surfaces. Reflectivity observations less than 10 dBZ were reassigned as clear-air observations and further thinned onto a 12 km grid. Radial velocity observations at these locations were discarded. The final processed radar observations were placed into 5-minute bins centered on each assimilation time and assimilated every 5 minutes. The observation errors for reflectivity and radial velocity were 5 dBZ and 2 m s^{-1}, respectively. It was determined that manual editing of the WSR-88D dataset to remove ground clutter, second trip echoes, etc. was unnecessary.

4.3.3 Assimilation details

The ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003) is used in the experiments. This algorithm is implemented in the DART software package and is the same algorithm used in Part I and described in section 2.2.6.

If the observations are assumed to have independent errors, then the data assimilation update can proceed serially, with the assimilation of each observation producing a new ensemble and forecast uncertainty estimate that is used as the prior ensemble for the next observation in the following update (this process is applied repeatedly until all observations are assimilated). Once the update is complete, each ensemble member is advanced to the next observation time using WRF.

As described in Part I, localization is needed to eliminate spurious covariances due to sampling error. Determining proper localization for surface observations is

challenging, especially in the vertical, where the coupling between surface observations and the overlying atmosphere is variable in time and space. Given the relatively dense observational network due to mesonet data across the portion of the domain where most convection is observed, a horizontal localization length scale that is somewhat smaller that was chosen in other studies is implemented here (60 km horizontal cutoff), using the Gaspari-Cohn localization function Gaspari and Cohn (1999). An 8 km vertical localization cutoff is used in the vertical. Horizontal and vertical localization for reflectivity and radial velocity observations was set to 12 km and 6 km, respectively, given the results of Part I. A 24 km (6 km) horizontal (vertical) localization cutoff was used for clear-air radar observations. The sensitivity of the analyses and forecasts to localization values is examined in a later section.

Two methods are used to counteract the tendency for spread values to decrease during the data assimilaton update. First, as in Part I, additive noise is applied in areas where reflectivity >25 dBZ immediately following the DA update and prior to the model advance (Dowell and Wicker 2009). This not only maintains spread, but accelerates the spin-up of convection. Second, adaptive prior inflation (Anderson 2009) is used to increase ensemble spread while preserving the ensemble mean. The adaptive inflation algorithm produces an inflation field that varies in time and space. The specfic parameters used in the additive noise and inflation algorithms, along with other DART settings, are provided in Table 4.3.

The full list of model fields within the DART state updated during the data assimilation procedure are provided in Table 4.4. Many of these fields are diagnosed (not prognosed) by WRF; these fields are incorporated within the DART state to make them available for the DART forward operators. For example, reflectivity and reflectivity-weighted fall speed are diagnosed within the microphysics scheme using assumptions consistent with those within the scheme. The reflectivity and

DART setting	**Value**
Filter type	EAKF
Ensemble members	50
Outlier Threshold	3.0
Adaptive Prior Inflation	Initial = 1.0, SD = 0.6
Adaptive Localization	1600
Localization type	Gaspari and Cohn (1999)
Localization (surface)	r_h = 60 km r_v = 8 km
Localization (radar, precip)	r_h = 12 km r_v = 6 km
Localization (radar, clear-air)	r_h = 24 km r_v = 6 km
Sampling error correction	False
Additive noise	0.5 m s^{-1}, 0.5 K

Table 4.3: Options for the DART assimilation system. Localization choices vary in some sensitivity experiments as described in the text. A full description of the adaptive inflation and additive noise algorithms are provided in Anderson (2009) and Dowell and Wicker (2009), respectively.

Model state fields updated during DA	
Horizontal wind (U and V)	Vertical velocity
Geopotential height	Potential temperature
Pressure	Water vapor mixing ratio
Cloud water mixing ratio	Rain water mixing ratio
Cloud ice mixing ratio	Snow mixing ratio
Graupel mixing ratio	Rain number concentration
Ice number concentration	Snow number concentration
Graupel number concentration	Condensational heating
10-m horizontal wind (U and V)	2-m temperature
2-m potential temperature	2-m water vapor mixing ratio
Surface pressure	Radar reflectivity
Reflectivity-weighted fall speed	

Table 4.4: List of fields updated during the data assimilation step by DART.

radial velocity forward operators directly interpolate these two fields during the data assimilation update to compute the predicted observations (i.e. $H(x^f)$).

4.3.4 Observation-space diagnostics and verification datasets

In Part I, analysis accuracy was assessed by comparing analyses directly to the true state. In the present experiments, the analyses are compared to observations before (prior) and after (posterior) assimilation. Thus, the state-space diagnostics (RMSE, bias, and CR in section 3.2.4) need to be redefined in terms of the observations.

It is helpful to define the observation *innovation*, d, as the difference between the observation and the model estimate of the observation[1]:

$$d^{a,f} = y^o - \overline{H(x_i^{a,f})} \tag{4.1}$$

Then, the previous diagnostics can be defined as:

$$RMSI^{a,f} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(d^{a,f})^2} \tag{4.2}$$

$$BIAS^{a,f} = \frac{1}{M}\sum_{i=1}^{M} d^{a,f} \tag{4.3}$$

$$Spread^{a,f} = \frac{1}{M}\sum_{i=1}^{M}[\frac{1}{N-1}\sum_{n=1}^{N}(H(x_{i,n}^{a,f}) - \overline{H(x_i^{a,f})})^2] \tag{4.4}$$

$$CR^{a,f} = \frac{(Spread^{a,f})^2}{(RMSI^{a,f})^2} \tag{4.5}$$

Here, M is the total number of observations at one assimilation time and H is the forward operator that computes the prior or posterior observation estimate using the model fields interpolated to the location of the observation. RMSE is re-defined as RMSI (root mean-squared innovation).

Surface and radar observations were also used to verify the forecasts at 5-minute intervals. The same sources of surface observations that were used during the assimilation period were used for verification. A three-dimensional gridded reflectivity mosaic from the NOAA National Mosaic and QPE system (Zhang et al. 2011) was used to verify model forecasts of composite reflectivity (CREF).

[1] d^f is called the innovation since the observation provides new information beyond what was predicted in the prior state (x^f). Although d^a is defined analogously to d^f, it is not traditionally referred to as an innovation.

4.4 Other convection-permitting ensemble forecasts

4.4.1 Description of other convection-permitting ensemble forecasts

Several operational and experimental convection-permitting, deterministic, and ensemble forecasts were available in real-time during this event. While the specifics of each modeling system and the source of forecast errors will not be described, their forecasts will be briefly summarized to gauge the ability of different forecast systems to provide accurate predictions for this convective event (the configuration can be found in the provided references). Much of the available experimental forecast guidance was generated in support of the 2012 SPC/HWT Spring Experiment (Kain et al. 2012). This included a 28-member, 36-hour convection-permitting ensemble forecast initialized daily at 00 UTC by the Center for Analysis and Prediction of Storms (CAPS) with 4-km horizontal grid spacing. A subset of 12 ensemble members possessed initial and boundary condition perturbations from the operational NCEP short-range ensemble forecasting system (SREF), while the other members varied in their use of physical parameterization packages. Two additional ensembles produced from operational high-resolution models (SSEO) and from the Air Force Weather Agency (AFWA) forecast system were also available.

Ensemble forecast neighborhood probabilities of updraft helicity (UH) in excess of $25\,m^2s^{-2}$ and ensemble maximum UH from the CAPS, SSEO, and AFWA forecast systems differ in their forecast of convection initiation and early evolution within the 21 UTC through 00 UTC time period (Fig. 4.10). The CAPS ensemble produces the largest probabilities (between 30-40%) in northwest Oklahoma (placed in the vicinity of CI1 and CI2) and western Arkansas (Fig. 4.10b). Coherent UH tracks are located along the OK/KS border and associated with more isolated convection

southward across the eastern TX Panhandle, southwest OK, and northern TX (Fig. 4.10a). Forecast probabilities less than 10% exist in CI3 (Fig. 4.10b). The SSEO had higher forecast probabilities (<50%) than the CAPS ensemble from CI1 eastward along the stationary boundary in southern KS, although forecast probabilities in the vicinity of CI3 were <10%, and only 1 or 2 ensemble members produced convection in this region (Fig. 4.10d). SSEO probabilities across western AR were also reduced compared to the CAPS ensemble. The AFWA ensemble produced large forecast probabilities (>60%) across south-central KS and northeast OK, east of CI1 and CI2 (Fig. 4.10f). This system also was the only of the three ensemble systems to produce extensive, long-lived convection, from southwest OK into central TX, mostly south of CI3.

In the 00 UTC through 04 UTC period, all three ensembles focused convection across northern and central OK, with similar biases present as in the previous 4-hour time period (Fig. 4.11). The CAPS ensemble produced relatively low probailities and was slow with the southward progression of observed convection, while forecast probabilities were larger in this region in the SSEO and AFWA ensembles. Across southwest OK and northern TX, one or two members from the CAPS and SSEO systems produce isolated, long-lived, supercells close to the track of the observed supercell that originated in CI3. Again, the AFWA ensemble contained a large amount of spread in the location of convection across north TX.

4.4.2 Discussion

Overall, the three ensemble forecasting systems showed some skill at predicting the general evolution of convection across this region, with errors typical of convective-scale predictions for the next-day convective cycle, in this case forecast lead times between 20-28 hours. Taken together, the three ensemble forecasting systems provide a sense of certainty in the development of a mesoscale environment supportive of deep

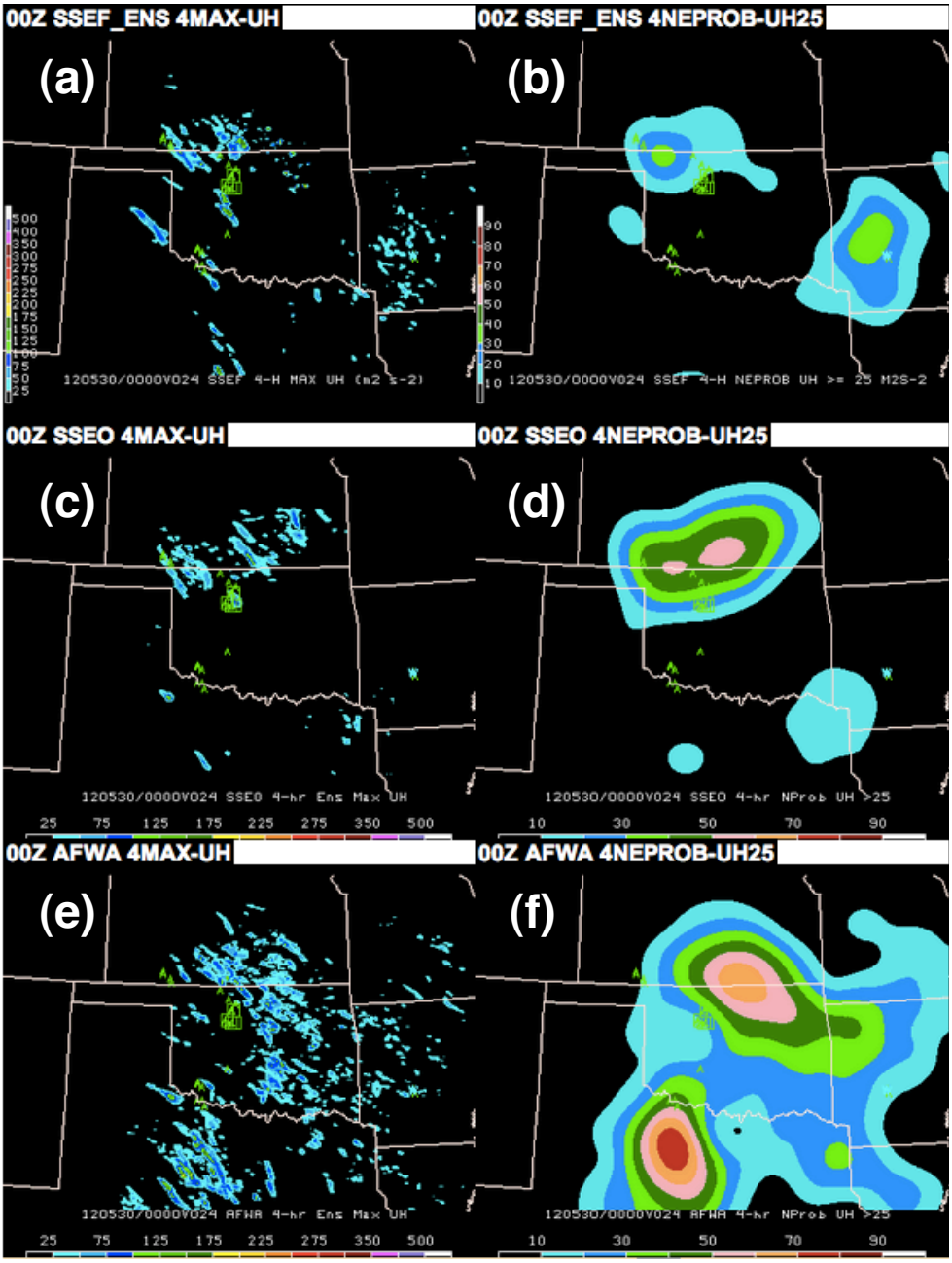

Figure 4.10: Forecasts of (a), (c), (e) ensemble maximum updraft helicity, and (b), (d), (f) ensemble probability of updraft helicity >25 m^2 s^{-2} from the (a), (b) SSEF, (c), (d) SSEO, and (e), (f) AFWA ensemble forecasting systems valid 20 UTC 29 May 2012 – 00 UTC 30 May 2012. Observed storm reports during the period are overlaid as green markers and squares.

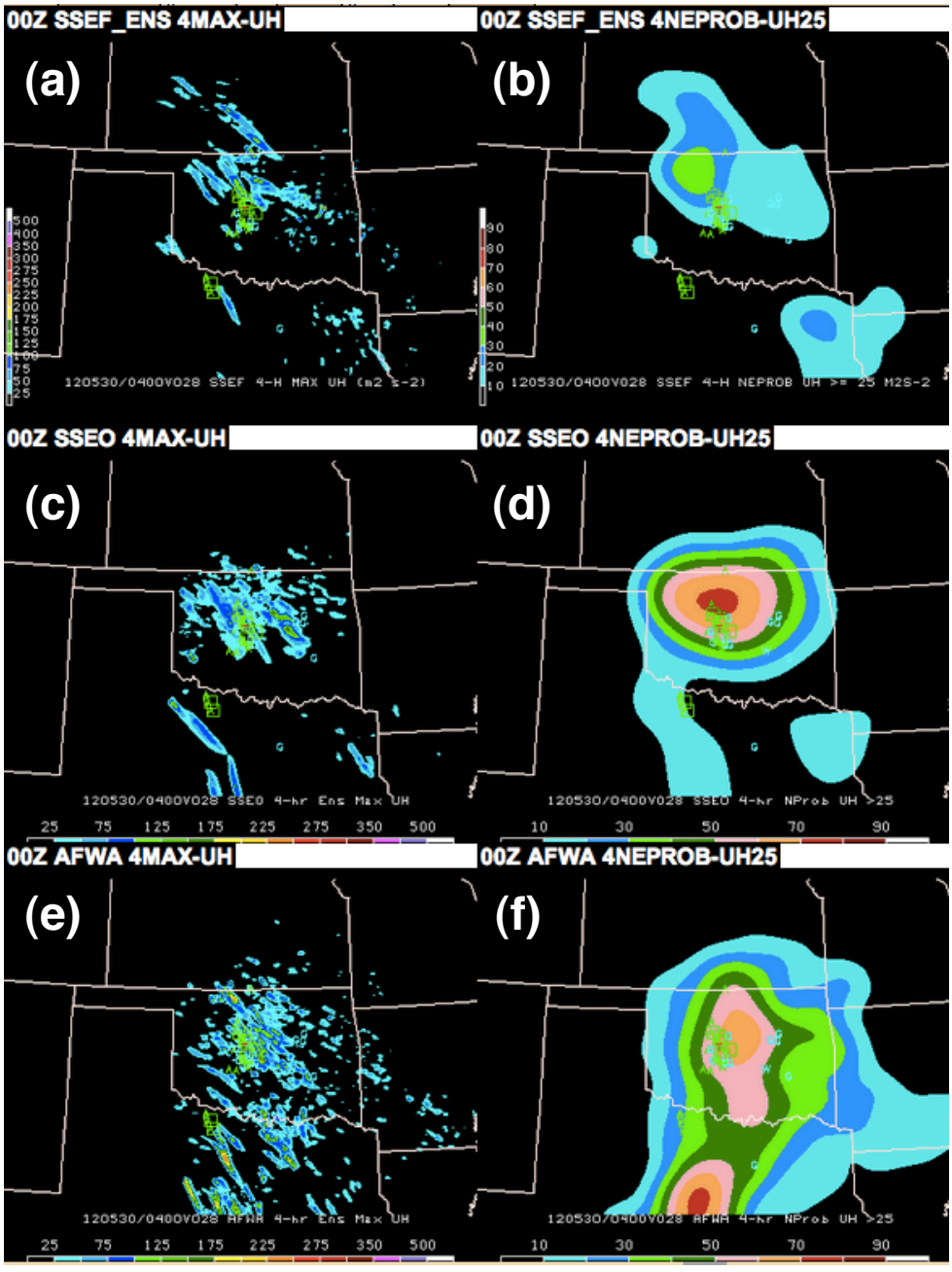

Figure 4.11: As in Fig. 4.10, but for the 00 UTC – 04 UTC 30 May 2012 forecast period.

convection across much of the region, including the placement of the salient surface boundaries tied to CI. Even in light of typical errors in the placement of convection at these lead times, interpreting this forecast guidance in real-time would lead to a fairly high degree of confidence in the following forecast scenario: initial convective development across northwest and northcentral OK that evolves toward the southeast into central OK. The forecast systems are less confident about the development and evolution of convection in southwest OK and northern TX. At least one ensemble member in each ensemble system produced a long-lived supercell near CI3. Thus, initiation within this region is less certain, and possibly less widespread, although if initiation occurs if appears likely that it will be sustained in an environment supporting of rotating deep convection.

In the assimilation experiments conducted herein, one of the goals is to improve forecasts of convection at lead times much shorter than the ensemble forecasts examined in this section. It is assumed that forecasts initiated at times closer to CI would be even more skillful than those presented here, but still retain convective-scale and meso-gamma scale errors that can potentially be ameliorated with data assimilation. Thus, the ensemble forecasts presented in this section lend confidence in the overall ability of contemporary ensemble systems to provide acceptably accurate forecasts for this case, thus not needing to correct for larger-scale errors in the initial conditions (e.g. boundary placement). Focus is thus directed toward improving smaller-scale details in the forecasts that will improve forecasts of convective initiation, upscale growth, and locations of severe weather. The next section will reinforce this idea by describing the ensemble forecasts produced as a control for the later data assimilation experiments.

4.5 Surface data assimilation experiments

4.5.1 Experiment design

Surface data observations were assimilated every 5-minutes between 18 UTC and 21 UTC. To assess the impact of various lengths of data assimilations, 50-member ensemble forecasts were launched at 19 UTC (SFC1H), 20 UTC (SFC2H), and 21 UTC (SFC3H). This encompasses CI and the early evolution of convection. These experiments are summarized in Figure 4.12. In addition to the three ensemble forecasts, a control ensemble forecast (CNTL) was initialized with the 18 UTC initial conditions from the downscaled NCAR mesoscale ensemble. No DA was used within this control forecast. In the following sections, the emphasis will be on the differences in the forecast of CI and subsequent convective evolution during the 21 UTC to 00 UTC forecast period between the various experiments.

4.5.2 Methods of forecast comparison

In both model output and observed radar data, convection was identified as areas where composite reflectivity (CREF) exceeded 25 dBZ (abbreviated as observed and forecast CREF25). This is a relatively liberal threshold, but was chosen to identify early signs of convective development and to identify convection in the ensemble mean, where CREF values may not exceed higher thresholds due to averaging among members, especially at longer forecast lead times. Several fields derived from the forecast CREF25 field were produced to evaluate differences between experiments and observations, including one-hour and three-hour ensemble probabilities of CREF25 (denoted PROB1H-CREF25 and PROB3H-CREF25, respectively). The PROB1H-CREF25 forecast was computed from the hourly-maximum CREF forecast, while the PROB3H-CREF25 forecast was computed from the three-hourly maximum CREF forecast derived with the individual hourly-maximum CREF fields over three

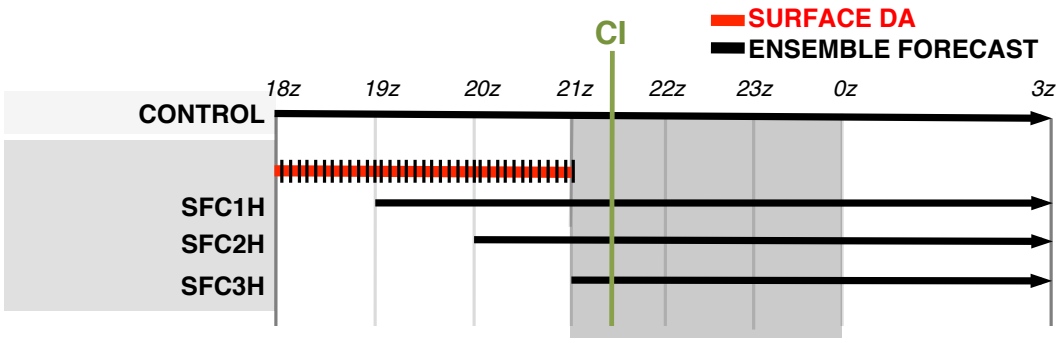

Figure 4.12: Summary of data assimilation experiments. Red lines indicate DA period, while black lines indicate 50-member ensemble forecasts initialized at different points during the DA period.

consecutive hours. At each grid point, the fraction of members where CREF exceeded 25 dBZ, using the one or three-hour maximum CREF field, determined PROB1H-CREF25 and PROB3H-CREF25.

In addition to the probabilities, two methods were developed to enable a more refined inspection of ensemble forecast CI timing errors. In the first method, areas where ensemble mean CREF exceeded 25 dBZ (denoted by MEAN-CREF25) were plotted at 15-minute intervals with earlier areas plotted on top of later areas. This provided a simple way to visually interpret ensemble mean timing errors. The second method used the 15-minute CREF output to compute the centroid of each CREF >40 dBZ object within each ensemble member. These points are color coded in time to provide a sense of the variation among ensemble members in timing and spread of convection. A size threshold of 10 grid points was used to remove small CREF objects. The same procedure was applied to the observed CREF field to produce a verifying centroid track.

4.5.3 CNTL ensemble forecast

In this section, the CNTL forecast will be summarized to provide some perspective on the performance of SFC1H, SFC2H, and SFC3H. Non-zero values of PROB3H-CREF25 are clustered near the areas of convective development in the three CI

regions. Within CI1, PROB3H-CREF25 values exceed 80% (Fig. 4.13a), with the spatial coverage of probabilities suggesting spread in the placement of convection. CNTL largely failed to initiate convection in CI2 during the three-hour period. Low values of PROB3H-CREF25 exist along the western edge of the observed CREF25 contour in CI2, associated with convection that develops late in the 21 UTC – 00 UTC period. The convective development in CI3 was well captured in CNTL, depicting two distinct locations of CI, with slightly larger probabilities for the southern storm. The extension of probabilities to the north and east with the northern supercell suggests the development of northeastward moving convection (i.e. left-moving supercells).

CNTL also produced spurious convection in several regions, the most notable being near CI4 in north central TX. Here, a focused area of higher PROB3H-CREF25 values (>65%) exists where a majority of ensemble members predicted the development of a pair of right-moving and left-moving convection, a scenario that was not observed, although visible satellite imagery indicated attempts at convection near 21 UTC in proximity to the PROB3H-CREF25 maximum (Fig. 4.4a). Other areas of spurious convection in CNTL exist along the stationary boundary in southeast KS and in association with terrain across western AR. While several isolated, short-lived areas of convection were observed in these locations, PROB3H-CREF25 from CNTL appear to overestimate convective coverage.

The PROB1H-CREF25 and MEAN-CREF25 fields both reveal timing differences between forecast and observed CI in CNTL. Within CI1 and CI2, no ensemble members forecast CREF25 within the 21 UTC – 22 UTC period, during the period when CI occurred (Fig. 4.14a). By 22 UTC to 23 UTC, a large area of low PROB1H-CREF25 values develops, centered on CI1 (Fig. 4.15a). This is also reflected in the MEAN-CREF25 field, with the earliest MEAN-CREF25 area occurring near 22:30 UTC (Fig. 4.17a). By 23 UTC, PROB1H-CREF25 magnitudes are largest near CI1 within the eastern half of the area where CREF25 was observed

(Fig. 4.16a). The low values of PROB1H-CREF25 along the western edge of CI2, noted in the previous section (Fig. 4.13a), develop within this hour. Thus, the ensemble members that do produce convection in this region initiate storms more than an hour later than observations.

In CI3, PROB1H-CREF25 values are (<20%) in the 21 UTC – 22 UTC period, in good agreement with the locations of initial development (Fig. 4.14a). MEAN-CREF25 areas are only associated with the southern storm in CI3, where PROB1H-CREF25 values are higher (this demonstrates the disadvantage of using ensemble mean forecasts of convection for low probability events; Fig. 4.17a). In CI3, non-zero PROB1H-CREF25 values expand and increase in magnitude during the following hour, with the southern supercell having larger probabilities (Fig. 4.15a). Later, the PROB1H-CREF25 envelope broadens within CI3, with some members producing convection west of the observed CREF25 area during this period, resulting in low probabilities (Fig. 4.16a). The lower PROB1H-CREF25 values in CI3 compared to the probabilities in CI1 imply uncertainty with the development of convection in CI3, potentially both with CI, and for those members that do produce convection, with the exact placement of the storms (i.e. larger spread).

4.5.4 Results from SFC1H, SFC2H, and SFC3H

The largest differences between the three ensemble forecasts occur within CI2 and CI3. For CI3, improvements are evident in the ensemble mean, as the development of the northern storm within CI3 is captured in the MEAN-CREF25 field in SFC2H and SFC3H (Fig. 4.17c-d), while it is not in CNTL and SFC1H (Fig. 4.17a-b). Second, as the DA period is increased, a westward shift occurs in the placement of CI within CI3, in better agreement with observations. Finally, the timing of CI is much improved, especially in SFC3H, with MEAN-CREF25 areas occurring as early as 21:30 UTC for the northern storm (Fig. 4.17d), while CNTL, SFC1H, and SFC2H delay initiation to

22 UTC or later in this region (Fig. 4.17a-c). Less substantial changes in the forecasts of MEAN-CREF25 occur for CI2. The MEAN-CREF25 area shifts westward in CI1 as the DA period is increased, but the timing of CI is further delayed, well after the time of observed CI shortly after 21 UTC (Fig. 4.17d). Very small areas of MEAN-CREF25 occur within CI2, the size and timing of which are nearly unchanged between the three experiments. The spurious convective development in CI4 is also unchanged across the set of four experiments.

Differences between the CNTL and SFC3H forecasts of PROB3H-CREF25 provide more insight into the differences in the MEAN-CREF25 forecast between the experiments (Fig. 4.13). In CI3, where the largest changes to MEAN-CREF25 occurred, PROB3H-CREF25 values are greater in SFC3H, especially associated with the northern storm (probabilities >80%). Here, an axis of higher probabilities (35-50%) exist associated with ensemble members that produce and maintain a left-split that moves toward the northeast. This was also evident in CNTL, but the PROB3H-CREF25 values are more focused, and more members predict this forecast scenario. While very small differences existed in the MEAN-CREF25 field within CI2 between the four experiments, the PROB3H-CREF25 values are larger in SFC3H than CNTL. The areas of non-zero PROB3H-CREF25 values in SFC3H coincide with the western edge of CI2, with between 35-50% of ensemble members indicating CI, including a right-moving supercell (not shown). The envelope of probabilities associated with CI1 is shifted westward and is more refined, matching well with the observed area of CREF25, although PROB3H-CREF25 magnitudes are reduced in SFC3H. In CI4, one area of CI was forecast in CNTL, but several additional areas are indicated in SFC3H, thus the surface DA appears to exacerbate the spurious convection within this region.

The forecast centroids suggest a large amount of uncertainty in the placement of convection within CI1 in both CNTL and SFC3H (Fig. 4.18). A large spread

of centroid points is present, although they envelope the observed centroid tracks. In SFC3H, some members produce convection earlier in the period (i.e. there are more blue centroid points in SFC3H than CNTL), and these are located further to the northwest in SFC3H. This gives a different impression of timing errors compared to the delayed CI suggested by the MEAN-CREF25 field. While PROB3H-CREF25 values were near zero in CNTL within CI2, the forecast centroid tracks show that several ensemble members produced convection at the end of the forecast period in this region, although the points are widely scattered and occur much later than observed. In SFC3H, a tightly clustered track of centroid points is forecast very close to the location and time of observed CI.

The centroids indicate a high degree of certainty in the placement and tracks of convection in CI3 in both CNTL and SFC3H (Fig. 4.18). In SFC3H, the number of forecast centroid points associated with the northern storm has increased and their spread has been reduced. Further, a larger number of members produce a northeastward moving storm associated with a left-split. Differences in the position and timing of the forecast centroid points for the southern storm are less compared to the northern storm between the two experiments. Within CI4, the forecast centroid points also indicate that the spurious PROB3H-CREF areas in SFC3H are associated with the development of three areas of convection; in CNTL, most members only predict spurious convection in one location.

4.5.5 Discussion

Taken together, the MEAN-CREF25, PROB3H-CREF, and forecast centroid tracks demonstrate that surface DA improved the forecasts of CI occurrence and timing within CI2 and CI3, while providing less benefit within CI1 and CI4. The CNTL ensemble was clearly deficient in producing convection within CI2 and had delayed CI in CI3. A potential reason for the delayed CI within CI3 in CNTL is the westward

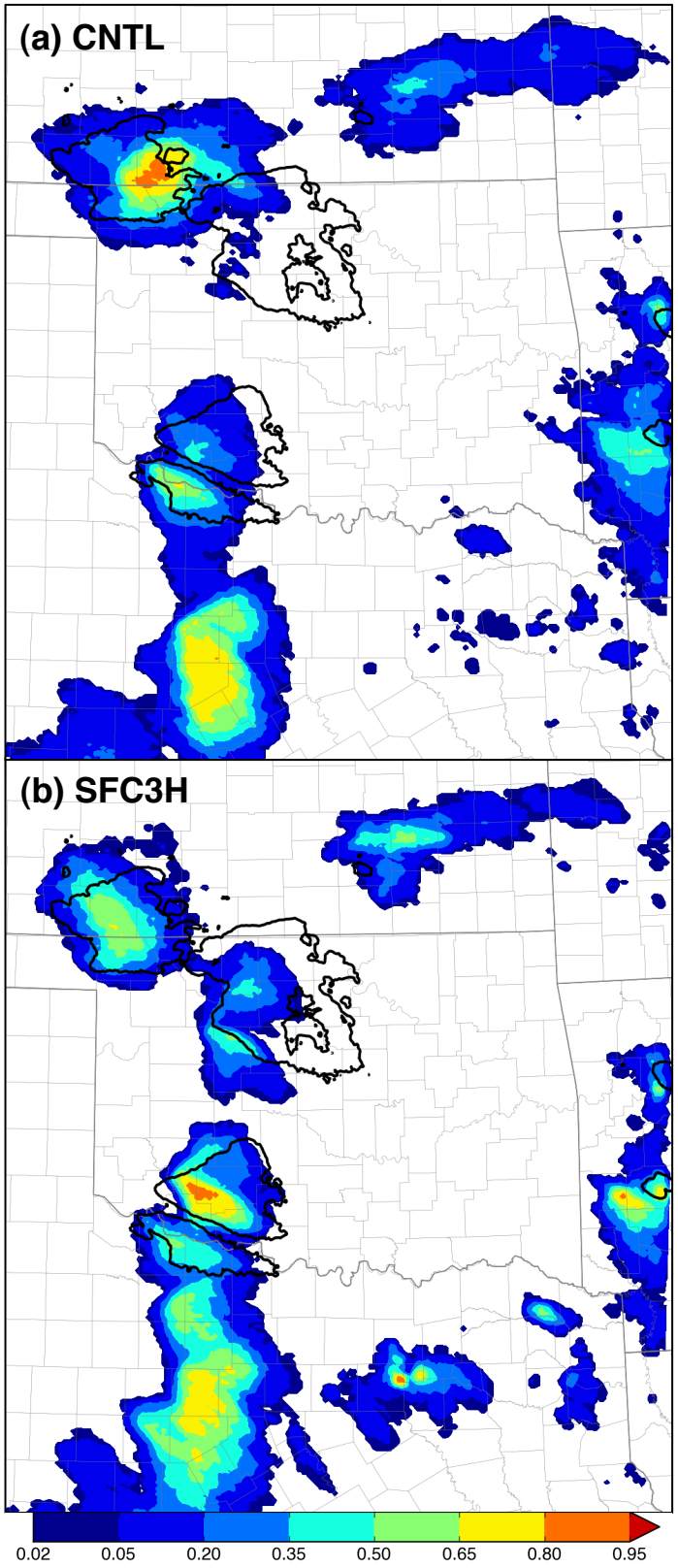

Figure 4.13: Three-hour (21 UTC to 00 UTC) forecast probabilities (shaded) of CREF >25 dBZ for (a) CNTL and (b) SFC3H. Maximum observed CREF >25 dBZ from 21 UTC to 00 UTC denoted by black contour.

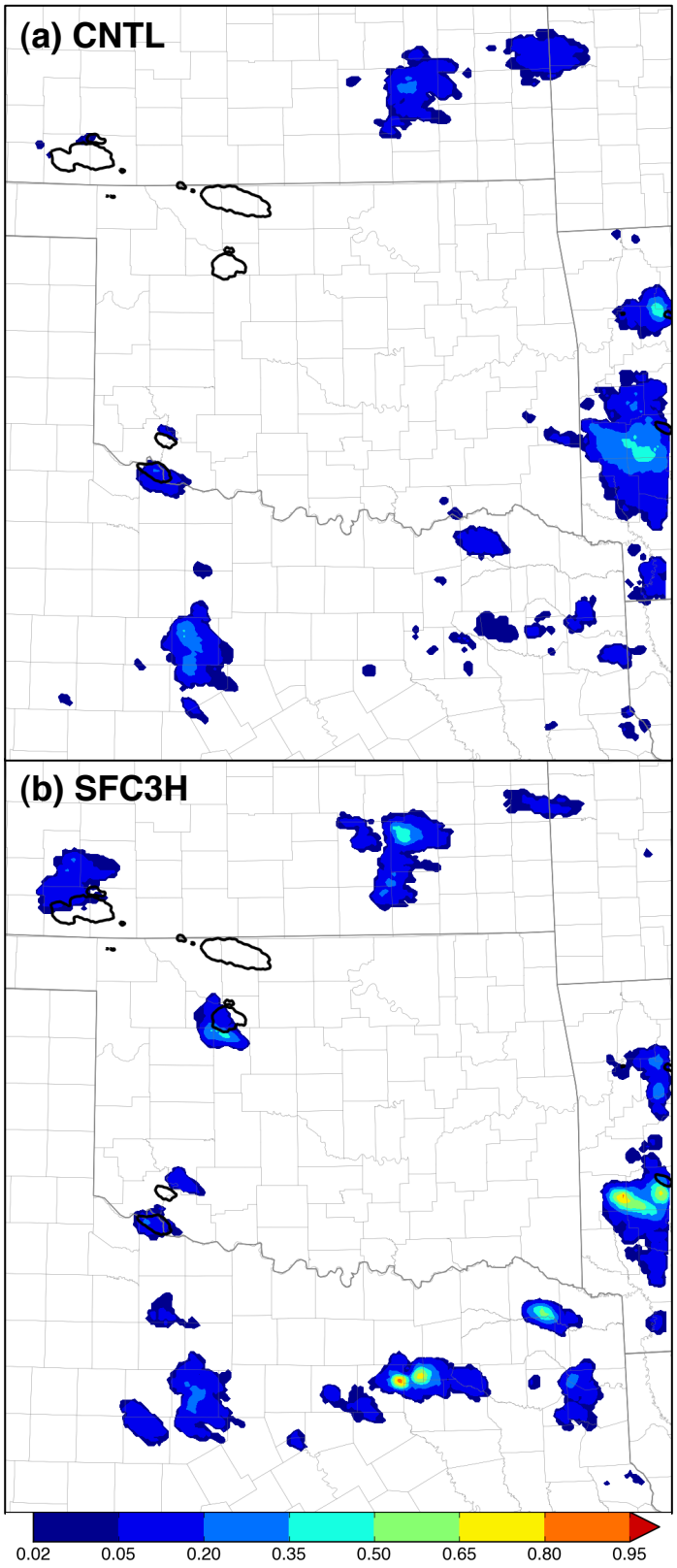

Figure 4.14: Same as Fig. 4.13 except for one-hour (21 UTC to 22 UTC) probabilities and maximum observed CREF contour.

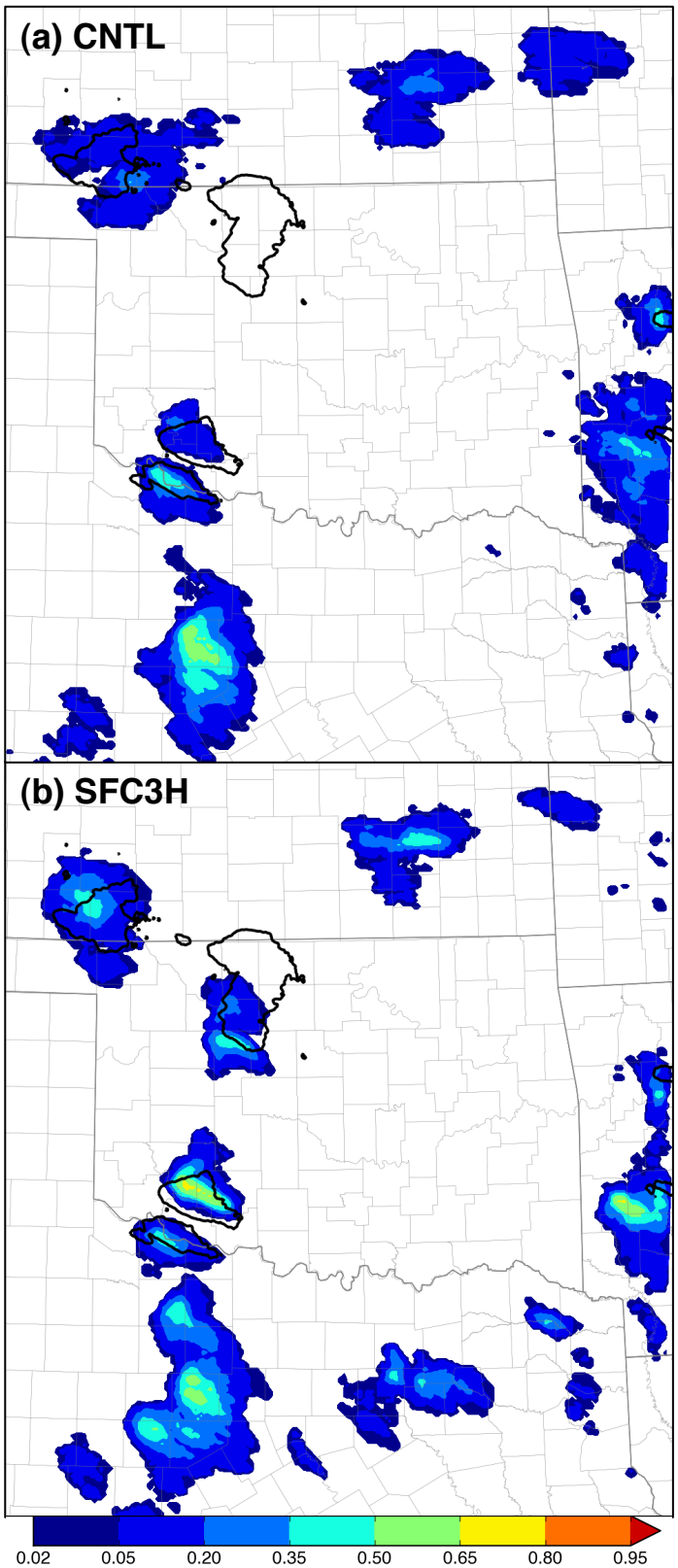

Figure 4.15: Same as Fig. 4.13 except for one-hour (22 UTC to 23 UTC) probabilities and maximum observed CREF contour.

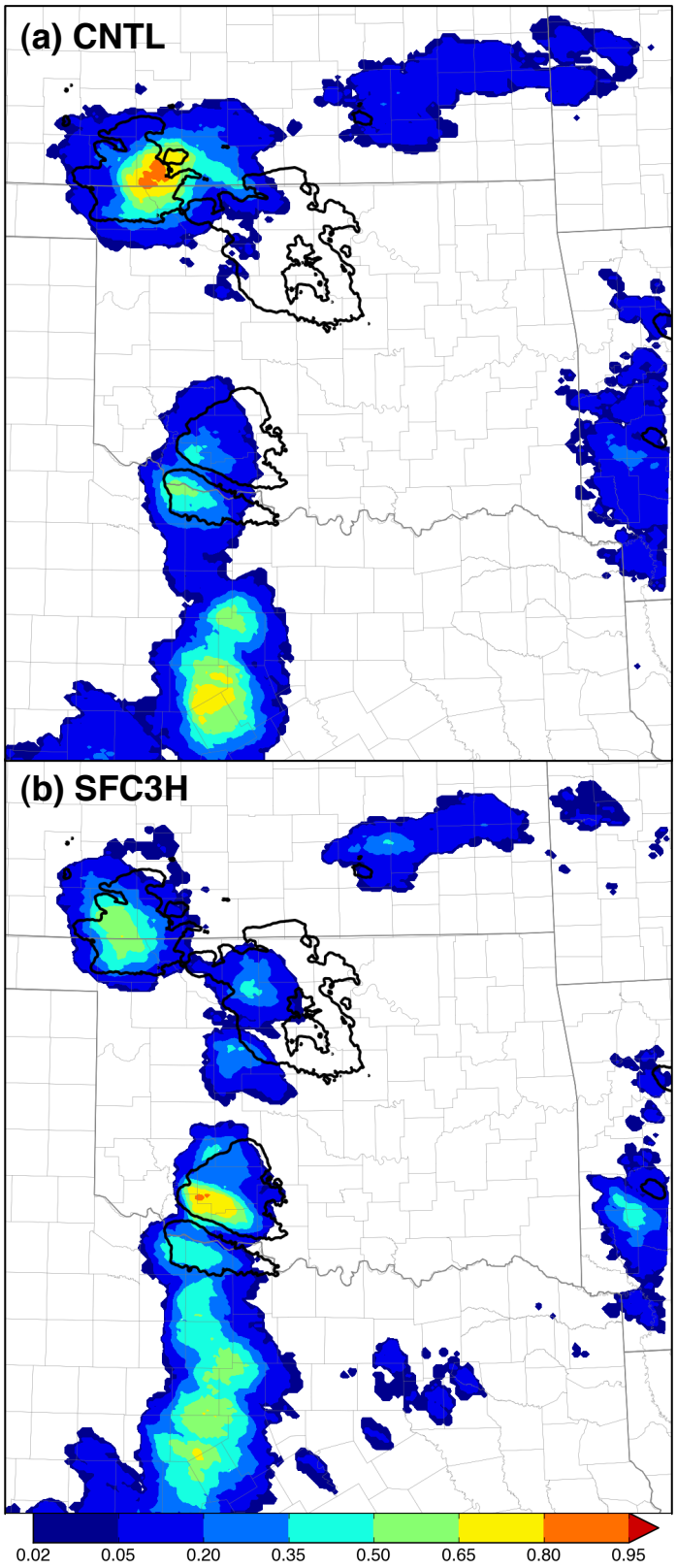

Figure 4.16: Same as Fig. 4.13 except for one-hour (22 UTC to 23 UTC) probabilities and maximum observed CREF contour.

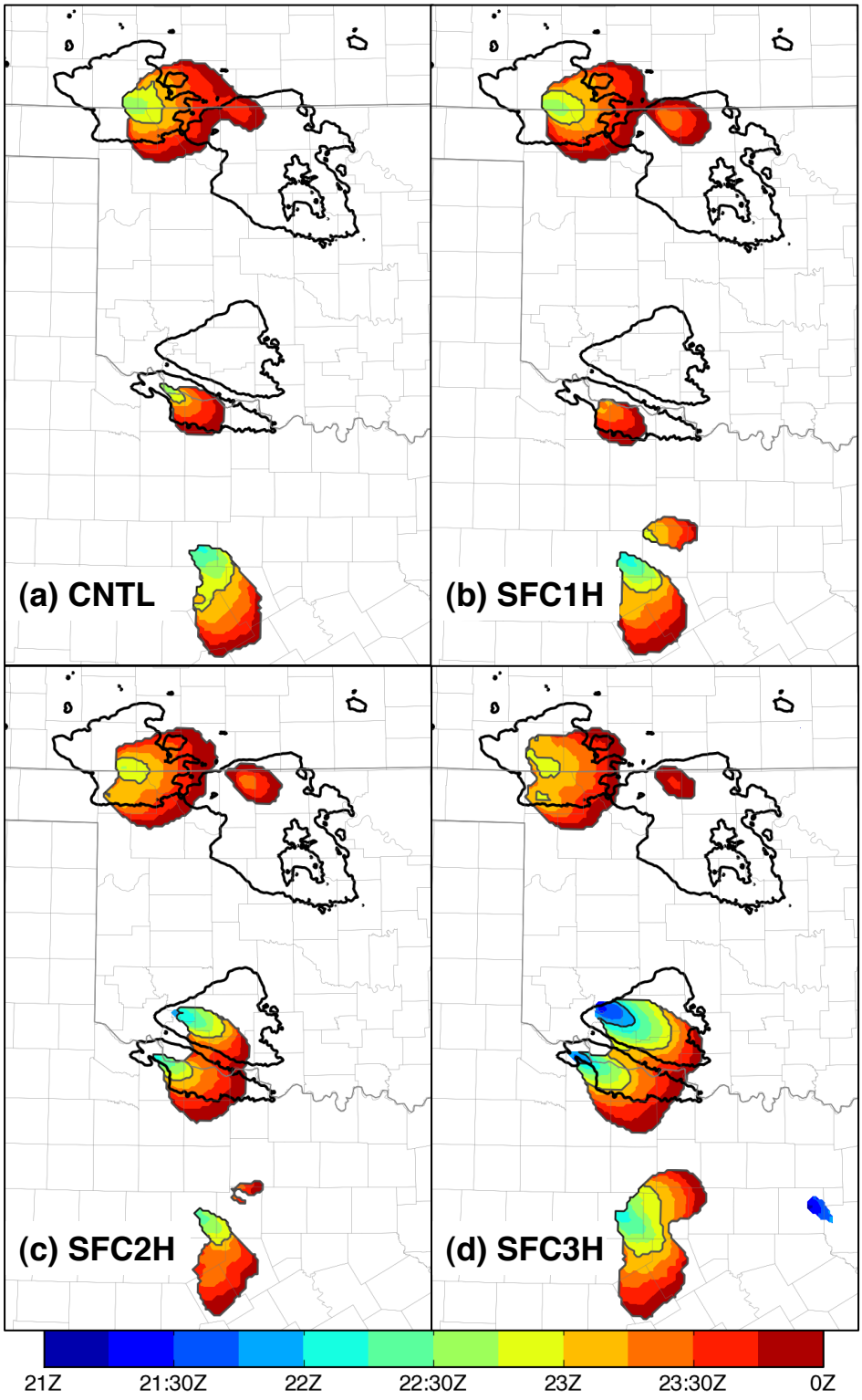

Figure 4.17: Ensemble mean CREF areas >25 dBZ color-coded at 15-minute intervals between 21 UTC and 00 UTC, with earlier areas plotted on top of later areas. Three-hourly maximum observed CREF >25 dBZ areas plotted with black contour.

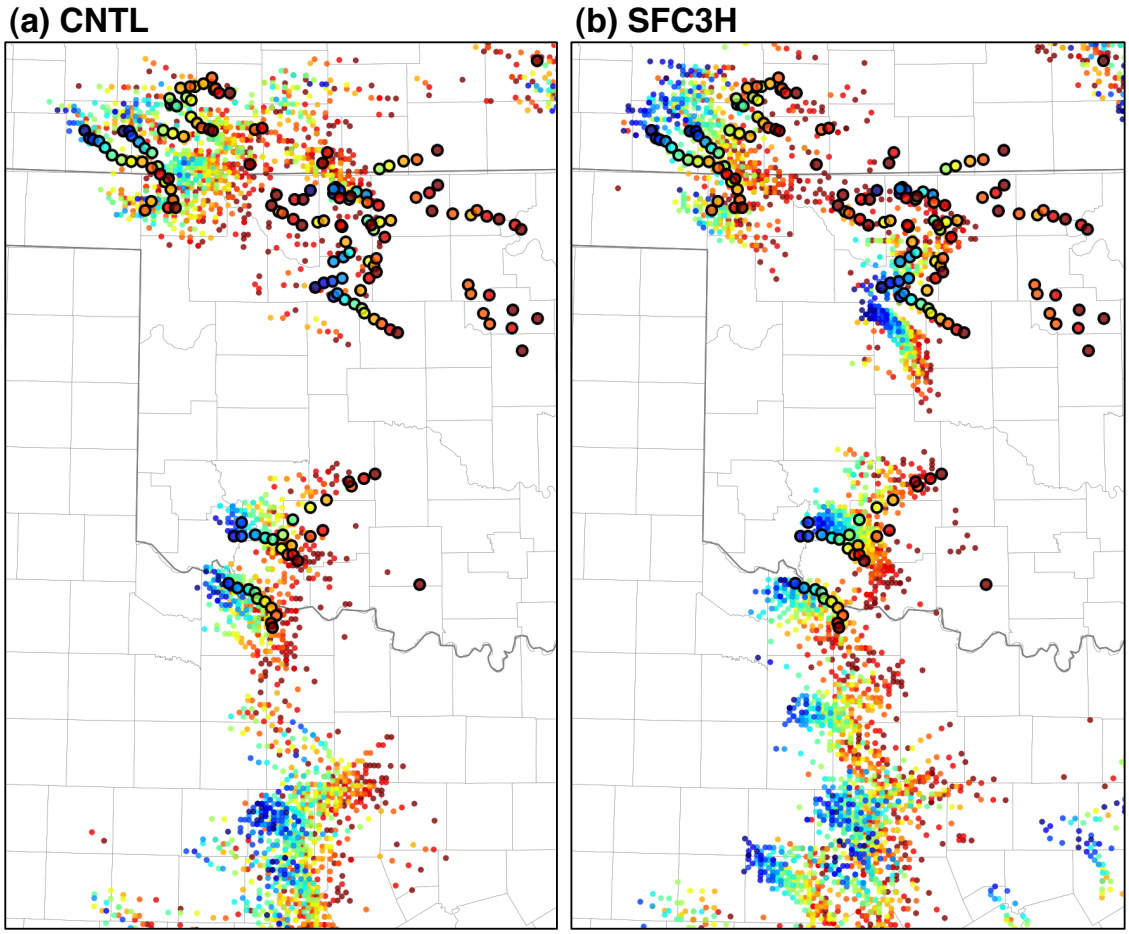

Figure 4.18: Centroids of forecast CREF25 objects from each ensemble member (colored circles) and observed CREF25 objects (colored circles with black outlines) at 15-minute intervals between 21 UTC and 00 UTC (colored circles). The creation process of the centroids are detailed in section 4.5.2.

bias in the placement of the surface dry line; CNTL does not mix the dry line far enough east during the afternoon hours. Other studies have found that convection-permitting models mix the dry line too far to the east (Clark et al. 2012). The present case is different in that the dryline is not associated with a synoptic-scale low-pressure system, thus its evolution is likely more sensitive to PBL estimates of vertical mixing. The dry line and moisture biases will be discussed in section 4.5.7.1.

Another potential reason forecasts were most improved in CI2 and CI3 are due to observation availability. CI2 and CI3 are relatively well observed (primarily by the Oklahoma mesonet), while fewer surface observations are available outside of Oklahoma within CI1 and CI4. The trends in CI timing and placement as the DA period increases are encouraging, especially within CI2 and CI3. The PROB3H-CREF25 magnitudes increase and become more refined as the time of CI approaches. In a real-world setting, probability trends between subsequent ensemble forecasts can give forecasters confidence in a particular forecast outcome. The precise effect of DA in these experiments is difficult to extract, since the forecast lead-time is not consistent. Thus, the differences in the above ensemble forecasts are likely due to a combination of changes in initial condition accuracy during DA and reduced predictability error with shorter lead times, making it difficult to disentangle each source of error on the resulting forecast. Two additional experiments were conducted in an attempt to understand these sources of error.

4.5.6 Impact of the length of the DA period

In a subsequent set of experiments, surface DA was delayed by one and two hours (assimilation period of 19 – 21 UTC and 20 – 21 UTC), so the end of the assimilation period coincides with 21 UTC. A cleaner comparison of the effects of DA length can be made with these two experiments (SFC1H-19UTC and SFC2H-20UTC), combined with SFC3H.

One hour of surface DA in SFC1H-20UTC has a clear impact on the PROB3H-CREF25 magnitudes compared to CNTL (c.f. Fig 4.13a, 4.19a). One hour of DA adjusts the envelope of probabilities westward within CI1, but magnitudes are reduced from >80% to <50%. Near and within CI2, two areas of >50% probabilities are produced, compared to near zero probabilities in CNTL. In CI3, PROB3H-CREF25 values increase by 40% – 50% compared to CNTL . A swath of probabilities is present in CI3 associated with the members producing a left-split. After additional hours of surface DA (Fig. 4.19b,c), PROB3H-CREF25 magnitudes increase within CI1, but are observed to decrease in other areas. In CI3 and CI4, PROB3H-CREF25 values remain similar to SFC1H-19UTC for the northern supercell within CI3, while PROB3H-CREF25 values decrease for the southern supercell. This is true for several areas of spurious convection within CI4 as well, although the probabilities increase associated with the spurious storm just to the south of the southern supercell in CI3.

The additional two experiments provide a clearer interpretation of the effect of multiple hours of surface DA on short-term forecasts of convection. In CI2 and CI3 probability values within SFC1H-20UTC are similar to those from SFC3H, thus it appears that in these two areas, the primary source of forecast improvement was the one-hour of surface DA between 20 UTC to 21 UTC, immediately prior to CI (Fig. 4.19a-c). Extending the DA period to two or three hours (as in SFC2H-19UTC and SFC3H) either has a small benefit, or is slightly detrimental to the PROB3H-CREF25 probabilities (e.g. reduction of magnitudes between SFC2H-19UTC and SFC3H). The longer period of DA is partially beneficial in CI3, since probability values increase between SFC1H-20UTC and SFC3H for the northern supercell, but decrease for the southern supercell. As noted before, a major role in these differences may be due to observational availability. In areas that are well observed (e.g. CI2 and CI3), the relatively high-density of observations, combined with frequent assimilation cycles, is able to adjust the state more quickly, thus reducing the need for a longer assimilation

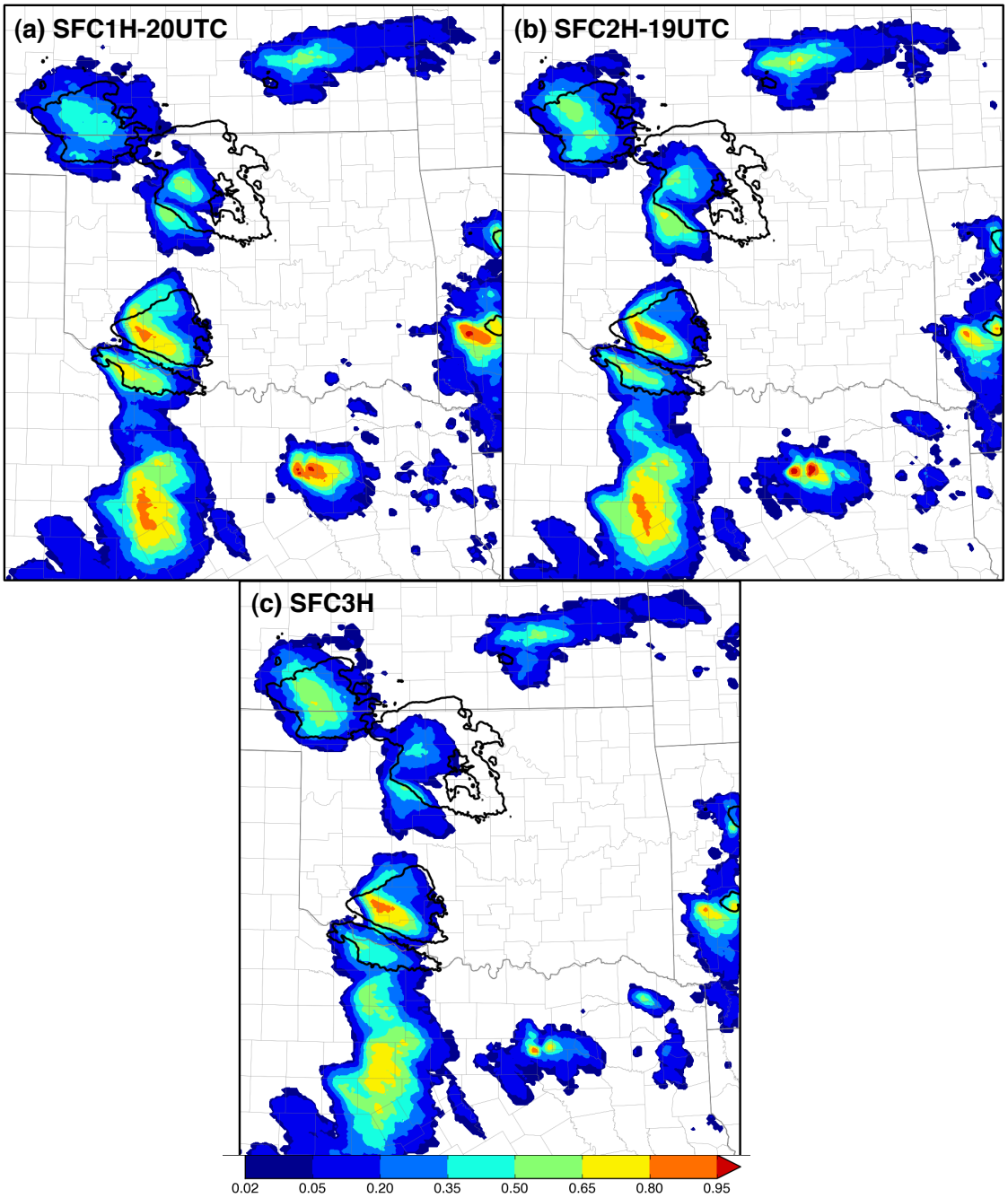

Figure 4.19: As in Fig. 4.13, but for (a) SFC1H-20UTC, (b) SFC2H-19UTC, and (c) SFC3H.

period. In more sparsely observed areas (e.g. CI1), the longer assimilation period is beneficial in order to reap improvements from better observed regions through forward integration of the model dynamics.

4.5.7 CNTL and SFC3H comparison

A comparison between the CNTL and SFC3H states is useful to understand the impacts and behavior of the surface DA process. Differences in the surface moisture field and in the structure of the lower troposphere will be highlighted in the following two subsections.

4.5.7.1 Differences in surface moisture field

CI can be especially sensitive to the distribution of boundary layer moisture (Crook 1996; Weckwerth 2000). Specific humidity variations of 1 g kg^{-1} can impact the occurrence of CI. Differences in the ensemble mean dew point field between CNTL and SFC3H were analyzed to gauge the impact of surface DA on the distribution of surface moisture (Fig. 4.20). The CNTL ensemble mean dew point forecast at 21 UTC possesses an approximately 4.8°C domain-average moist bias (Fig. 4.20a), with the largest errors within the domain located in two regions. The first is associated with the surface dry line, where observations indicate that the placement of the dryline in the forecast is too far west. The second is an axis that stretches from southcentral KS into southeast OK, near and to the northeast of a weak stationary boundary. The three hours of surface DA in SFC3H reduces both of these surface moisture biases (domain-average moist bias of 2.2°C). The largest differences between CNTL and SFC3H are in these two regions (Fig. 4.21).

The improvements in CI in these areas between CNTL and SFC3H are likely driven by these differences in the surface moisture field and the simulated dry line circulation. Not only is the dry line placement improved in SFC3H, but the magnitude

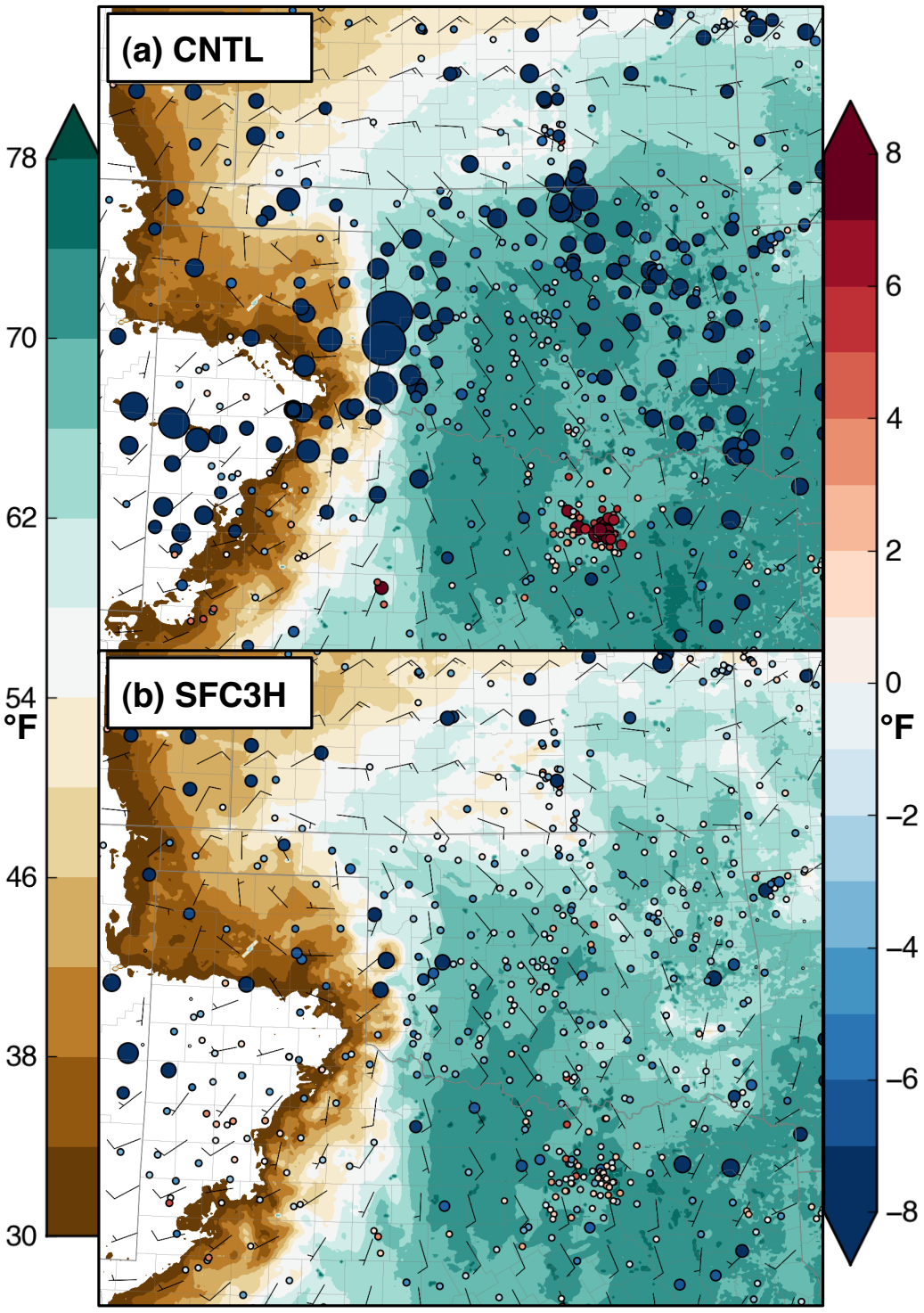

Figure 4.20: Ensemble mean 2-m dew point temperature (shaded; degrees Fahrenheit) and 10-m wind field at 21 UTC 29 May 2012 from (a) CNTL and (b) SFC3H. Filled circles are differences between the ensemble mean dew point analysis and observations (forecast minus observed).

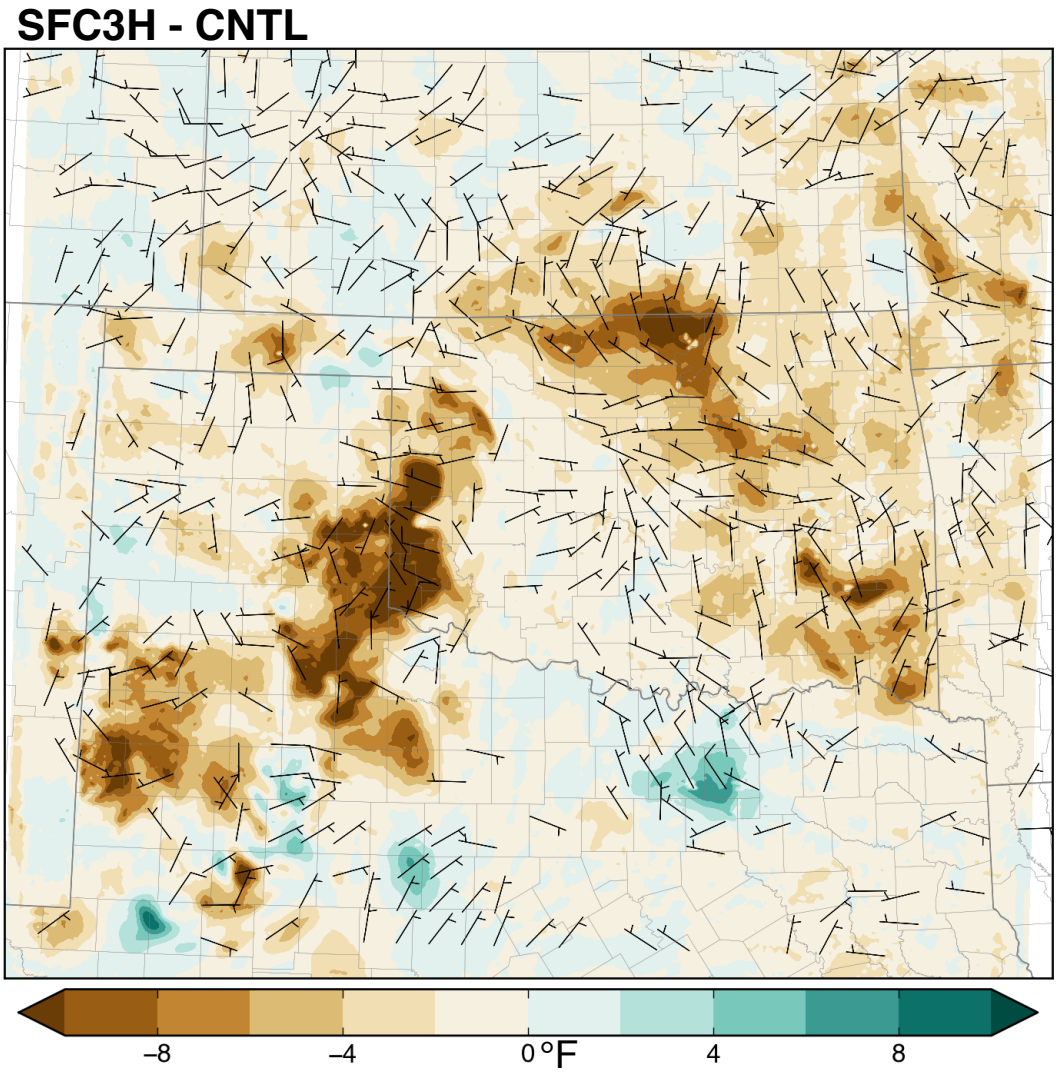

Figure 4.21: Difference field (SFC3H-CNTL) for the ensemble mean 2-m dew point temperature (shaded; degrees Fahrenheit) and 10-m wind field at 21 UTC 29 May 2012. Full wind barbs indicate a wind difference of 5 m s^{-1}, while half barbs is a difference of 2.5 m s^{-1}.

of the differences between the theta-e of the two airmasses is also larger, due to lower dew points behind the dryline across SW OK and NW OK and little to no change in the surface dew point ahead of the dryline in western OK. The increased theta-e difference in SFC3H is associated with a stronger dryline circulation and surface mass convergence in the areas where convection was observed to initiate. This is further reflected in the ensemble mean surface wind field differences between the two experiments in the vicinity of the surface dry line. Three hours of surface DA resulted in increased westerly momentum behind the dryline and easterly momentum ahead of the dry line (Fig. 4.21).

4.5.7.2 Differences in vertical profiles

Surface DA has the ability to also make adjustments to the state above the surface. To illustrate the changes in vertical profiles between CNTL and SFC3H, vertical temperature and dew point profiles from each of the 50 posterior ensemble members at 23 UTC at the KOUN (Fig. 4.22) and KAMA (Fig. 4.23) sounding locations were plotted on a skew-T diagram. These two stations are representative of the air mass ahead of and behind the surface dry line. The observed KOUN and KAMA radiosondes, launched at 23 UTC, are overlaid for comparison. The vertical profiles are only shown below 400 hPa; differences between the two experiments are small above this level, since surface observations can only modify the state below 8 km AGL due to vertical localization.

The 23 UTC CNTL ensemble mean temperature profile agrees well with the KOUN observed sounding, with the ensemble showing very little spread throughout the profile. The largest temperature spread is in association with the height of the PBL. The PBL height in the ensemble mean is less than observed, with the sounding falling within the spread of the ensemble members. All of the members are too moist within the boundary layer and the observed profile falls well outside the envelope

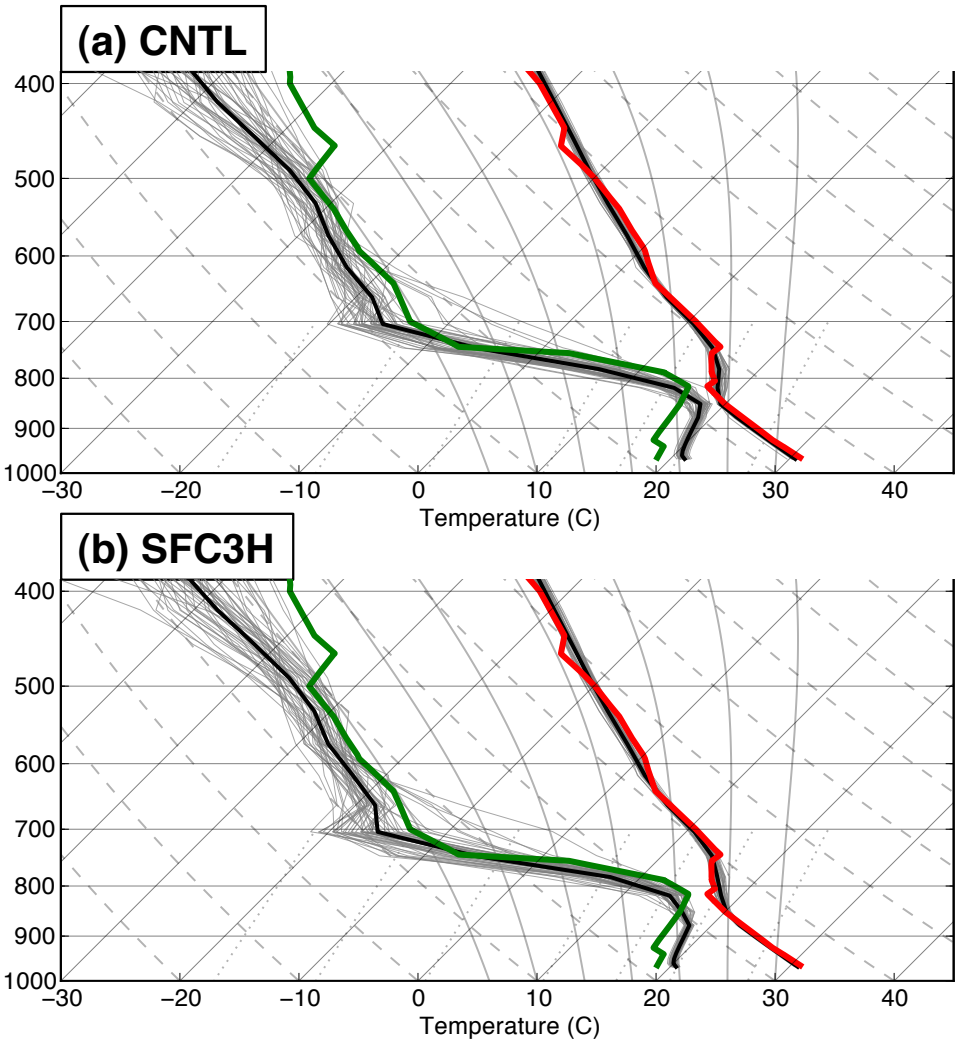

Figure 4.22: Skew-T temperature (degrees Celsius) and dew point (degrees Celsius) profiles from ensemble members (gray) and ensemble mean (black) at 23 UTC 29 May 2012 from (a) CNTL and (b) SFC3H at KOUN. Observed temperature (red) and dew point (green) profiles from the 00 UTC 30 May 2012 KOUN radiosonde (launched at 23 UTC).

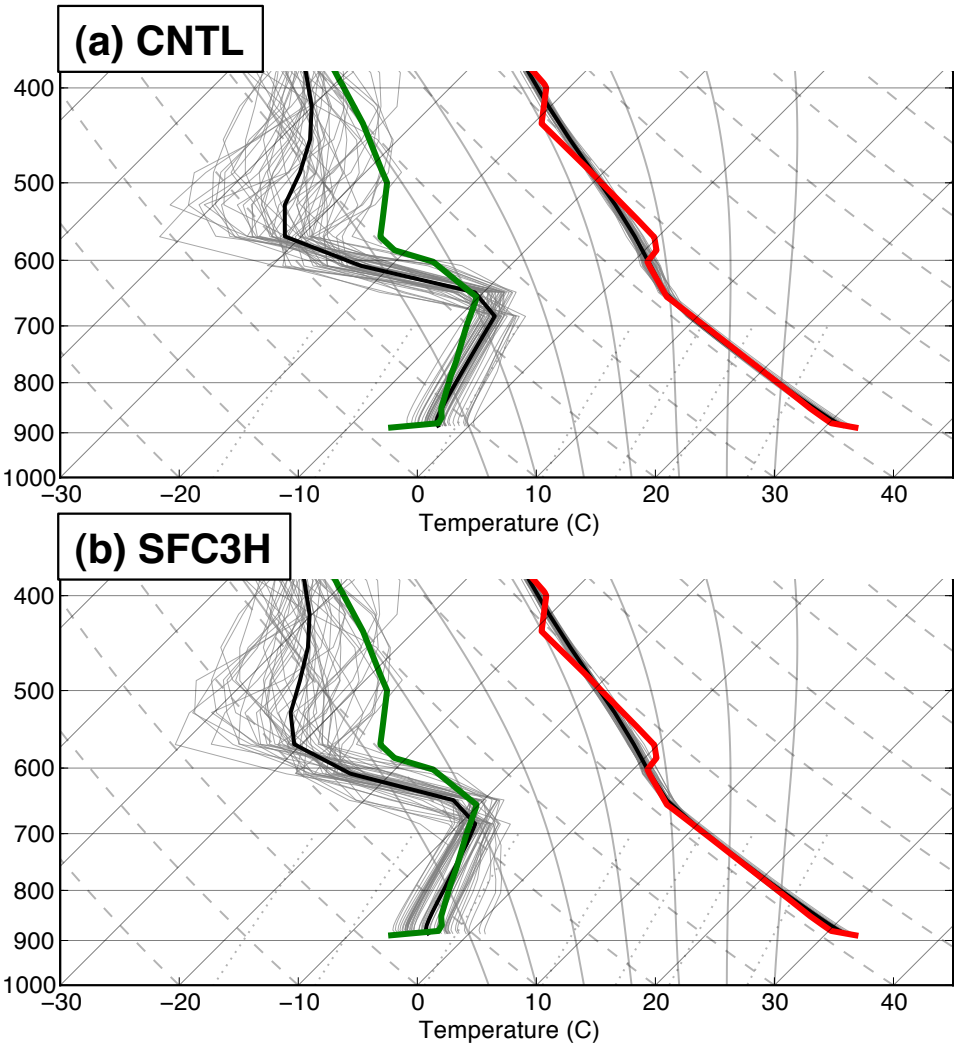

Figure 4.23: Same as Fig. 4.22, but for the ensemble forecast and observed profile at KAMA.

of the CNTL members. This is consistent with the moisture biases discussed in the previous section. Above the PBL, slightly better correspondence exists between the observed and CNTL ensemble mean within the inversion layer (850 - 750 hPa).

Surface DA is unable to make significant changes to the temperature profile due to little ensemble spread in temperature (i.e. the ensemble gives little weight to observations when the ensemble spread is small relative to the observation error). The SFC3H mean moisture profile is approximately 0.5 g kg^{-1} less than the CNTL mean profile within the PBL, closer to the observed sounding. Yet, the boundary layer depth has been reduced in SFC3H compared to CNTL (neither experiment captured the inversion at the PBL top). Similar adjustments to the moisture profiles occur at KAMA. The moisture amount has been reduced within the PBL, although the height of the PBL has not deepened considerably, and has been reduced in many of the ensemble members.

The 23 UTC wind profile in both CNTL and SFC3H is 10-15 knots slower than was observed at KOUN below 3 km AGL (Fig. 4.24). It appears that surface DA is unable to correct this lower-tropospheric wind bias since CNTL and SFC3H have similar errors that are inherited from the mesoscale background. Also, both CNTL and SFC3H possess a slow bias in wind speed at KOUN throughout the troposphere. Discussion about the effects of these errors on predicted supercell storm motion is provided in Part III, since this bias is expected to have the largest impact later in the forecast period as convection matures.

4.5.8 Impact of mesonet data and frequent cycling

Two additional assimilation experiments were conducted to assess the impacts of both the mesonet dataset and the 5-minute assimilation interval on the analyses. In the first experiment, mesonet data were withheld; the remaining METAR observations were assimilated at 5-minute intervals from 18 UTC – 21 UTC (SFC3H-NOMESO). In a

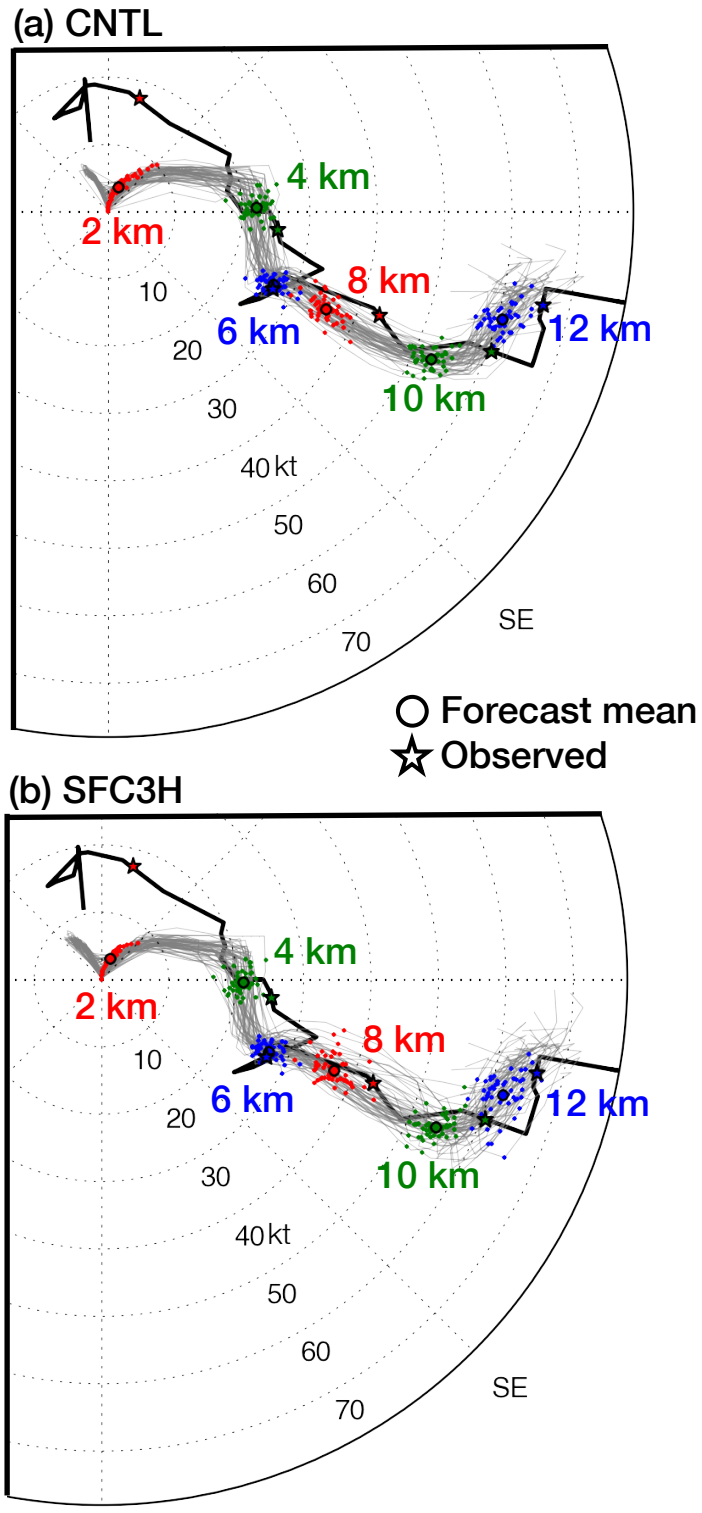

Figure 4.24: Hodographs from (a) CNTL and (b) SFC3H ensemble members (gray) at 23 UTC 29 May 2012 and observed 00 UTC 30 May 2012 hodograph (black) at KOUN. Points on each hodograph indicate the wind speed/direction (kts) at 2 km, 4 km, 6 km, 8 km, 10 km, and 12 km MSL for each ensemble member (colored circles), the ensemble mean (colored circle with black outline), and the observed hodograph (colored stars).

second experiment, observations were assimilated once per hour (SFC3H-HOURLY). In SFC3H-HOURLY, METAR observations taken between 15-minutes before and 15-minutes after the top of the hour were assimilated, along with mesonet observations within 2.5 minutes of the top of the hour. The smaller mesonet assimilation window in SFC3H-HOURLY was chosen to ensure that each mesonet observing site only contributed, at most, one observation per assimilation cycle. SFC3H-HOURLY was designed to mimic hourly, cycled, DA systems that have been used in previous studies (e.g. Wheatley et al. 2012).

The 21 UTC ensemble mean surface dew point and wind fields in SFC3H-NOMESO and SFC3H-HOURLY share more similarities to CNTL than to those from SFC3H (Fig. 4.25a-b). The largest adjustments in SFC3H-HOURLY are made in association with the dry line in SW OK and NW TX, extending into west-central TX (Fig. 4.25a). Even smaller differences are present between SFC3H-NOMESO and CNTL (Fig. 4.25b). Given that neither experiment is capable of reducing the moisture biases nor fully adjusting the position and magnitude of the dry line, it appears the combination of frequent assimilation cycles and higher-resolution mesonet data that provides the largest impact on the analyses.

To determine if the adjustments made by the surface DA persist into the free forecast period, the ensemble mean dew point RMSE was computed every 5-minutes using METAR and mesonet observations for the four experiments (CNTL, SFC3H, SFC3H-NOMESO, SFC3H-HOURLY; Fig. 4.26). The initial differences in error between the experiments at the initial analysis time continue throughout the forecast. Following a period of fairly rapid error growth during the first 30 minutes, the RMSE for all experiments stabilize. The SFC3H ensemble mean has the lowest RMSE throughout the 3-hour period. CNTL and SFC3H-NOMESO are largely indistinguishable during the entire period, except for the first 15-minutes when SFC3H-NOMESO has a slightly smaller RMSE than CNTL. The RMSE for

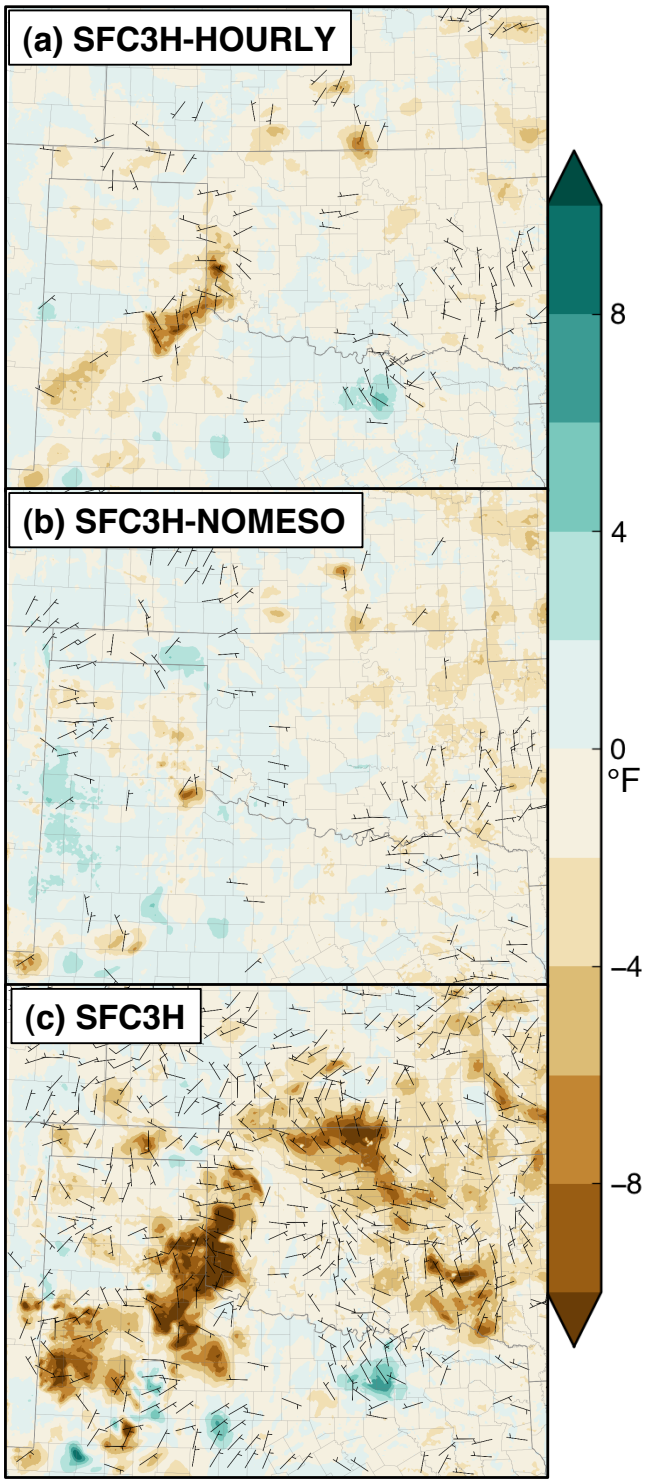

Figure 4.25: As in Fig. 4.21, but for (a) SFC3H-HOURLY minus CNTL, (b) SFC3H-NOMESO minus CNTL, and (c) SFC3H minus CNTL.

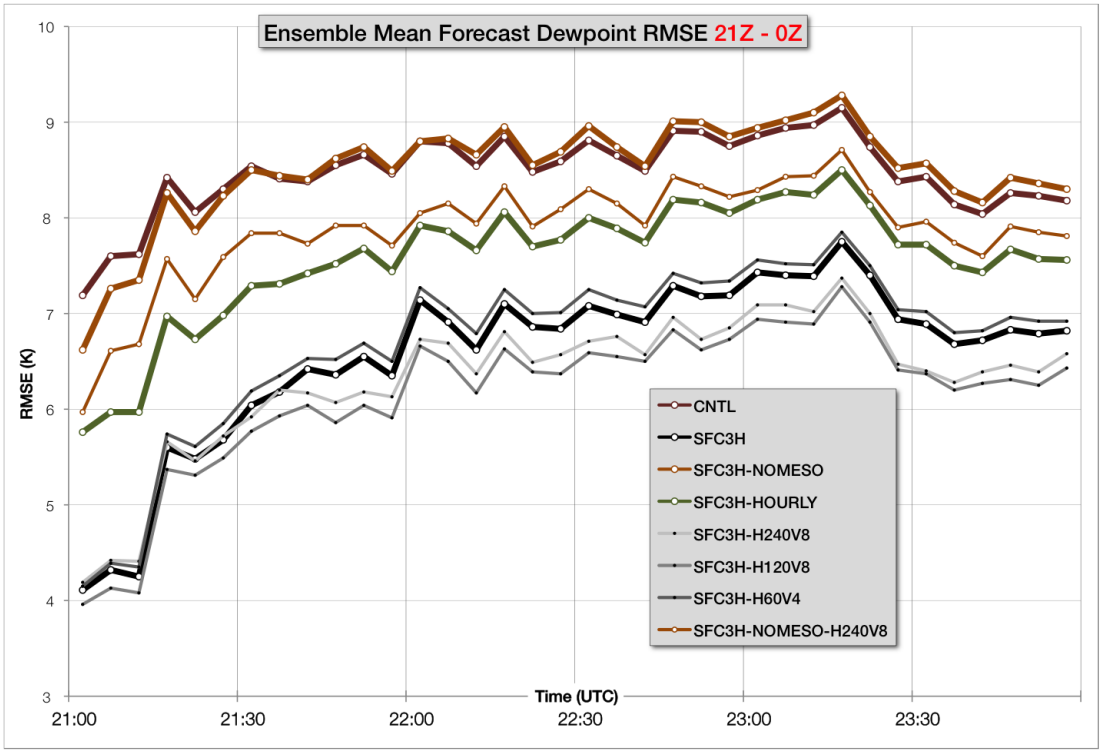

Figure 4.26: Root-mean squared dew point temperature error (K) for the 21 UTC to 00 UTC ensemble mean forecast from various surface DA experiments.

the SFC3H-HOURLY falls approximately midway between SFC3H and CNTL during the period. Assimilating METAR and mesonet observations once per hour is able to cut the dew point errors in half in this case, with an equally large amount of RMSE reduction if the observations are assimilated every 5-minutes. In the latter case, the benefits are primarily due to the ability to utilize the mesonet data that occur beyond the restrictive mesonet data window used herein. It is uncertain what effect broadening the mesonet assimilation window would have on the analyses, to include observations further away from the analysis time. Doing so would frequently result in multiple observations, valid at the same location but at different times, being assimilated in an analysis valid at a single time. This would likely have the largest effect where observations are rapidly changing, for example in the vicinity of fronts and, in the present case, a surface dry line.

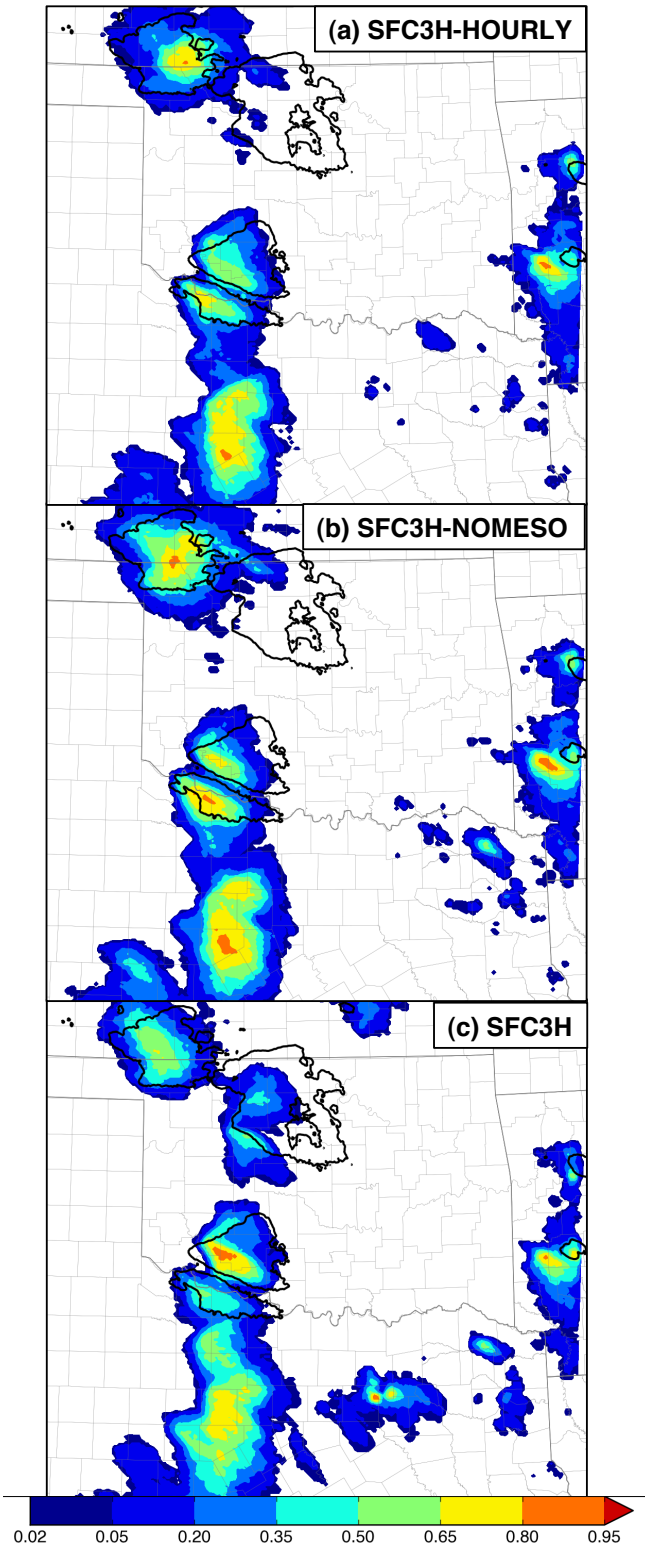

Figure 4.27: As in Fig. 4.13, but for (a) SFC3H-HOURLY, (b) SFC3H-NOMESO, and (c) SFC3H.

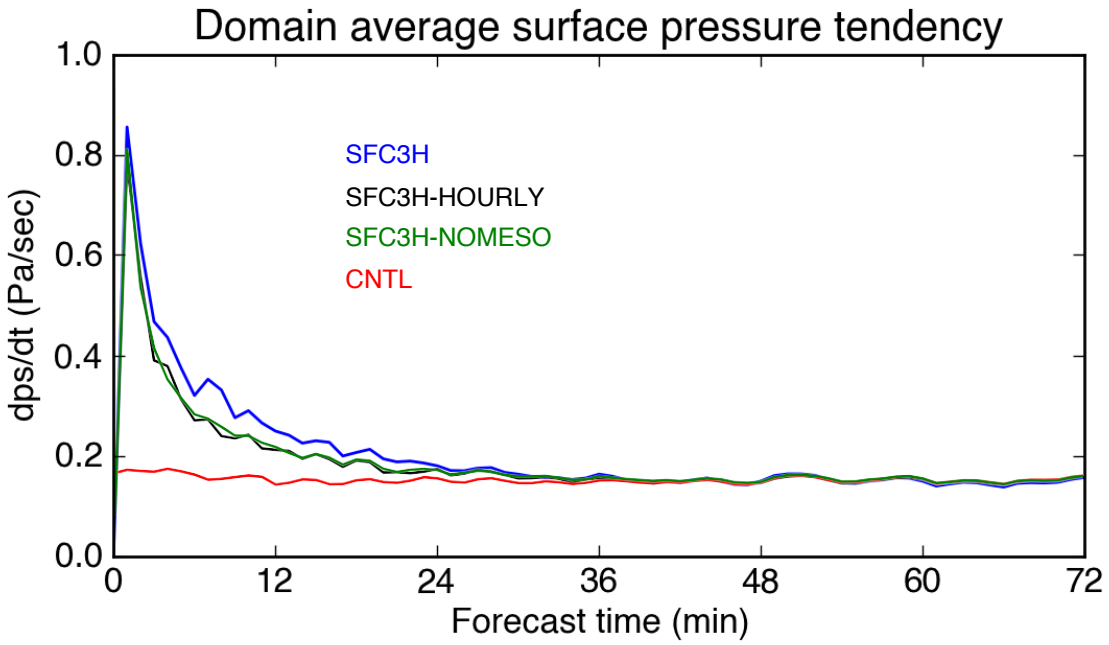

Figure 4.28: Ensemble-averaged, domain-averaged, absolute surface pressure tendency (Pa sec^{-1}) computed at 5-minute intervals for a 6-hr forecast beginning at 21 UTC from SFC3H (blue), SFC3H-HOURLY (black), SFC3H-NOMESO (green), and CNTL (red).

PROB3H-CREF25 values for the two experiments generally fall in between CTRL and SFC3H (Fig. 4.27). One notable exception is the probabilities associated with CI2; in this area both SFC3H-HOURLY or SFC3H-NOMESO are similar to the CTRL in that they produce little to no convection. Only SFC3H has ensemble members that develop sustained convection within CI2.

Assimilating surface observations as frequently as every 5-minutes could lead to imbalance in the resulting analyses, potentially to the detriment of forecast accuracy (Greybush et al. 2011). The surface pressure tendency (PSFCTEND) provides a way to measure the degree of imbalance in the initial analyses. Thus, the ensemble mean domain maximum PSFCTEND (that is, the average of the 50 values of domain maximum PSFCTEND) was computed at 5-minute intervals in through the forecast period for the ensemble forecasts from the three assimilation experiments (SFC3H, SFC3H-HOURLY, and SFC3H-NOMESO). Only the absolute magnitude of the PSFCTEND was used in the computation.

The peak in surface pressure tendency is similar across the three experiments, with domain average PSFCTEND between 0.75 and 0.85 Pa s^{-1} (Fig. 4.28). For SFC3H, PSFCTEND decreases more gradually than SFC3H-HOURLY and SFC3H-NOMESO. By 2 hours into the forecast, the experiments have indistinguishable PSFCTEND values. Thus, imbalance is slightly larger in SFC3H and takes longer to stabilize than the other two experiments, even though SFC3H-NOMESO also uses a 5-minute assimilation frequency (although with a much smaller number of assimilated observations).

4.5.9 Sensitivity to horizontal localization cutoff

As described in Part I, the choice of an appropriate length scale for localization is a complex function of observation type, density, state variable, etc. Some results are presented here to illustrate differences in the surface dew point and forecasts of convection when larger horizontal localization values are used. The inclusion of mesonet data in the current study results in a varying observation density across the domain (Fig. 4.8). The horizontal localization cutoff in all of the experiments so far (60 km) was chosen to be approximately double the distance of the average observation spacing within Oklahoma, the region of relatively high-density observations due to the presence of the Oklahoma mesonet. More sparsely observed regions might benefit from a larger horizontal cutoff. To test this hypothesis, SFC3H was repeated using a horizontal localization cutoff increased to 120 km (SFC3H-H120V8) and 240 km (SFC3H-H240V8).

The differences in the ensemble mean dew point RMSE between the three experiments at the final analysis time (21 UTC) is relatively modest (approximately $0.3°C$). Compared to SFC3H, the RMSE is smaller in SFC3H-H120V8, while SFC3H-H240V8 has a slightly larger RMSE. Larger differences emerge later in the forecast period, primarily after 21:30 UTC. Both SFC3H-H120V8 and SFC3H-H240V8 have

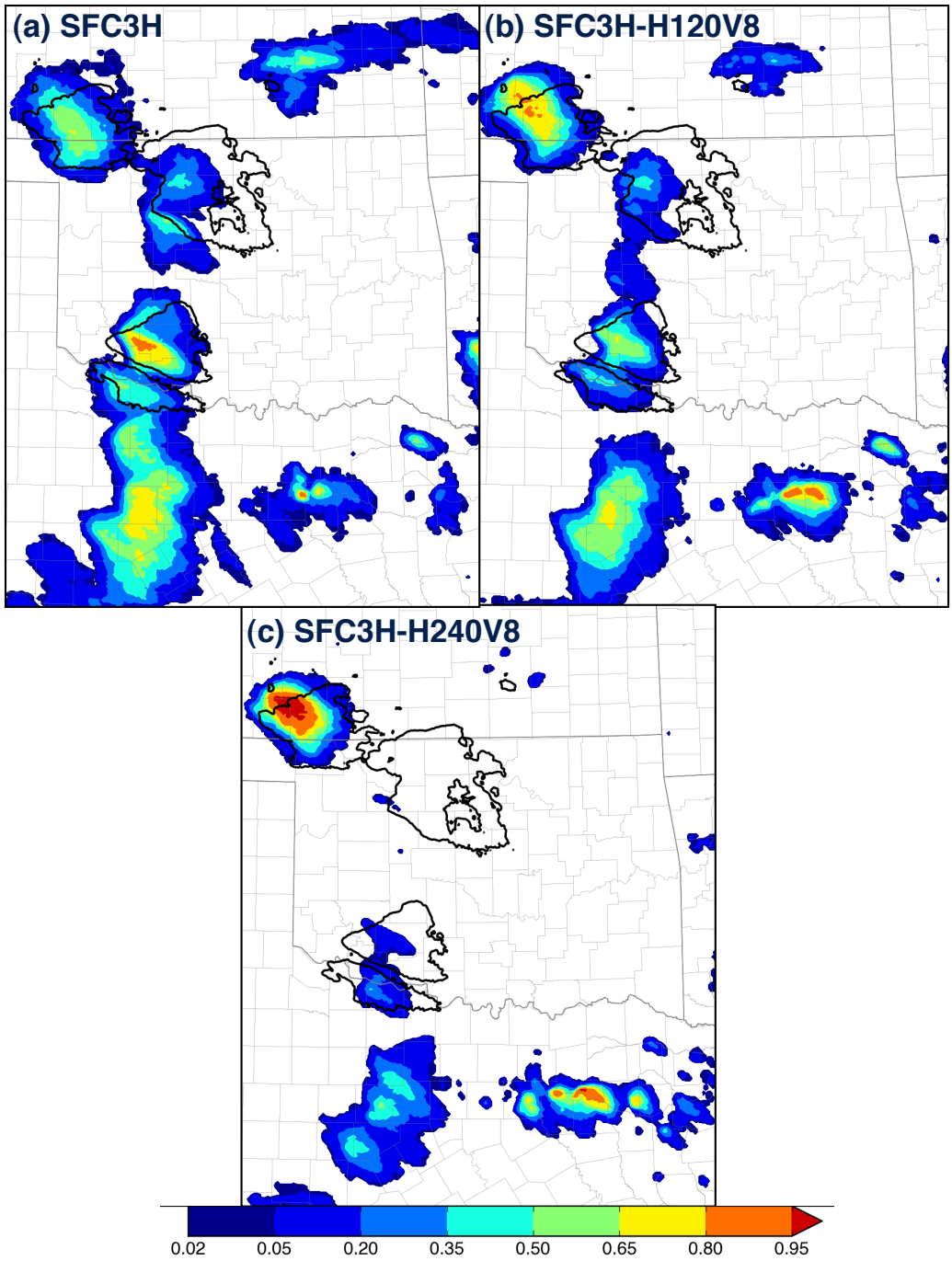

Figure 4.29: As in Fig. 4.13, but for (a) SFC3H, (b) SFC3H-H120V8, and (c) SFC3H-H240V8.

smaller ensemble mean dew point RMSE than SFC3H, with SFC3H-H120V8 the smallest of the three experiments. Increasing the localization length results in a smoother analysis, which is likely the reason for the lower RMSE in SFC3H-H120V8, but increases to the localization beyond this length-scale result in slight RMSE increases in SFC3H-H240V8, although the RMSE remains less than SFC3H for the forecast period beyond 21:30 UTC.

The forecast RMSE implies that the SFC3H-H120V8 and SFC3H-H240V8 analyses both have better representations of the surface dew point field, yet these benefits do not necessarily extend into better short-term forecasts of convection. In fact, fairly large differences exist between the three experiments. In general, increasing the horizontal localization cutoff reduces probabilities of convection throughout the domain in the 21 UTC – 00 UTC period (Fig. 4.29). One exception to this is in CI1, where PROB3H-CREF25 magnitudes are increased. In CI2, forecast probabilities are reduced to near zero in SFC3H-H240V8, while in CI3 a similar reduction occurs, although <35% of the ensemble continues to produce two areas of convection, albeit with much larger placement errors than in SFC3H or SFC3H-H120V8. In CI4, the reduction in forecast probabilities is beneficial since no convection is observed in this region. All three experiments continue to produce large values of PROB3H-CREF25 east of CI4 near KDFW; forecast probabilities are less sensitive here to localization than in other areas of the domain.

Two factors are likely influencing the differences in forecast probabilities between these three experiments: 1) spread reduction and 2) smoother analyses, both due to larger localization. As localization increases, the ensemble spread is reduced as state points are influenced by more observations (this effect was described in Part I). Over time, this spread collapse could result in observations having little weight on the analyses (i.e. filter divergence) as the ensemble becomes increasingly confident on a given forecast outcome. Unlike the experiments in Part I, the use of prior

inflation in these experiments is designed to counteract this tendency, by increasing the prior ensemble spread. If prior state inflation were insufficient to ameliorate the tendency for spread collapse, probabilities of convection would likely be drawn to lower or higher values, as members cluster around solutions where convection develops or is suppressed. This behavior is observed between the three experiments, with CI1 CREF25 probabilities increasing (almost all members produce convection in this location), and probabilities decreasing in areas along the dry line in OK and TX.

Larger localization also tends to produce smoother analyses as state points are influenced by more observations. In the present experiments, this effect may reduce the magnitude of circulations along the dry line, the primary initiating boundary within the domain. Probabilities are indeed reduced along the dry line within OK and TX, although the increase in probabilities in SW KS cannot be explained by this effect. Given the reduced observation density within SW KS compared to nearby areas in OK, the larger localization may be improving the analyses in this region by incorporating information from far away observations, a potentially more substantial effect than that of reducing the circulations associated with surface boundaries. If this were occurring, a similar effect would expect to be seen in other areas of limited observations, e.g. along the dry line in central TX; this is not observed to occur in this region.

4.5.10 Summary and Discussion

Results from several surface DA experiments were discussed in the present chapter. In the primary experiment, SFC3H, surface data, including METAR and mesonet data, were assimilated every 5-minutes during a three-hour period extending from 18 UTC and ending at 21 UTC, shortly before CI. High-frequency assimilation of surface data led to improvements in forecasts of CI, both in the timing and placement of

initial convective development. These improvements extended into the short-term forecasting period due to a more accurate representation of the surface moisture field, and reduction of errors due to model biases. Additional experiments that only assimilated routine surface observations (i.e. METAR), or only assimilated observations each hour, did not see the same improvements in the forecasts of CI. Finally, increases to the horizontal localization cutoff for the assimilated surface observations led to detrimental forecasts of CI for this case.

Frequent assimilation of surface observations, including routinely issued METARs and observations from mesoscale networks, improved forecasts of CI and subsequent convective evolution where the development of convection is driven largely by surface boundaries (e.g. a dry line). This is the first time that high-frequency mesoscale surface data have been assimilated, with a rapid cycling period using the EnKF, for the explicit goal of predicting CI for a real convective event. Given the challenges associated with forecasting CI, this is an encouraging result, yet several challenges remain that were also documented within this section. Those involve model error from PBL parameterizations and the effects that surface observations have on the overlying free atmosphere.

The MYJ PBL parameterization scheme used within these experiments produced forecasts with a positive moisture bias, as noted in previous evaluations (Hu et al. 2010). A similarly configured mesoscale ensemble system to the one used herein also contained a positive moisture bias within the PBL that impacted forecasts of convection (Romine et al. 2013). In cycled DA systems, like the system that provided the initial and boundary conditions for the present experiments, these model errors can persist and grow with each analysis cycle, unless observations are regularly assimilated to constrain the model solution. It appears the 5-min cycling frequency used herein is able to reduce the moisture bias, in addition to correcting the placement and strength of surface boundaries and their associated surface circulations.

Even though surface DA improves the analyses, assimilating surface observations with biased PBL parameterizations may still be suboptimal (Dee and Da Silva 1998). Future work should investigate the use of techniques to reduce the effects of model errors, particularly PBL parameterization error, on meso- and convective-scale EnKF analyses.

Appropriately adjusting the state above the surface during surface DA is an additional challenge to the successful assimilation of real surface observations. While surface fields may be improved, the state above the surface may not necessarily be more accurate. In the current experiments, model error is again a likely culprit. While a suboptimal update due to the surface moisture biases likely plays a role in these errors, examination of the soundings suggests that errors within the PBL parameterization that describe mixing, entrainment, and PBL growth may have led to erroneous covariances between observations and the model state. For example, at both KOUN and KAMA, the boundary layer was shallower and moister than the observed profile in the CNTL ensemble without surface DA. Thus, after a period of surface DA, it would be encouraging to see a deeper PBL, along with a corresponding reduction in mixing ratio. Although the SFC3H mean PBL mixing ratio was indeed less, the PBL depth became shallower, instead of the desired deepening. Further research is crucial to fully understand the role of model error due to PBL parameterization and to begin to develop strategies to properly handle surface observations and their impact above the surface. One simple strategy is to reduce the vertical localization length-scale, thus removing the impact of surface observations above a specified height.

The improvements to forecasts of CI and evolution in most of the domain by assimilating surface observations is taken to be a sign that the mesoscale environment is more accurately represented in the final ensemble analyses after multiple hours of surface DA. These benefits should extend to analyses and forecasts that are derived from experiments where radar data are assimilated once they become available

following CI. In Part II, the impact of radar DA compared to surface DA for this case will be investigated.

4.6 Surface and radar data assimilation experiments

Radar data were assimilated for two hours (21 UTC to 23 UTC) following the three hours of surface DA. Surface DA is also performed during this period. Reflectivity, radial velocity, and clear-air reflectivity observations were assimilated. The analyses and forecasts from this experiment (SFCRAD5H) will be analyzed in the following section.

4.6.1 Observation-space diagnostics

4.6.1.1 Domain-averaged statistics

Observation-space diagnostics were computed to assess the fidelity of the assimilation of radar observations, particularly observations of radar reflectivity that have been assimilated with varying degrees of success in previous work. Within this section, diagnostics are computing using observations that were successfully assimilated (and not those that were rejected). During the two hours of radar DA, radial velocity observations are fairly well-behaved (Fig. 4.30a). The posterior RMSI quickly stabilizes to just under 2 m s^{-1}, slightly larger than the prescribed observation error standard deviation. The prior total spread and prior RMSI behave consistently, especially by the end of the assimilation period, resulting in prior consistency ratios near 1. After the first 30 minutes of DA, bias (i.e. mean innovation) values approach zero, although after 22:15 UTC, bias values become slightly positive. This occurs during a large increase in the number of assimilated observations, and will be investigated section 4.6.2.

Reflectivity RMSI gradually decreases during the first hour of assimilation, with the prior RMSI stabilizing to just below 6 dBZ by 22:30 UTC (Fig. 4.30b). The more gradual fit to reflectivity observations, and a larger RMSI, compared to radial

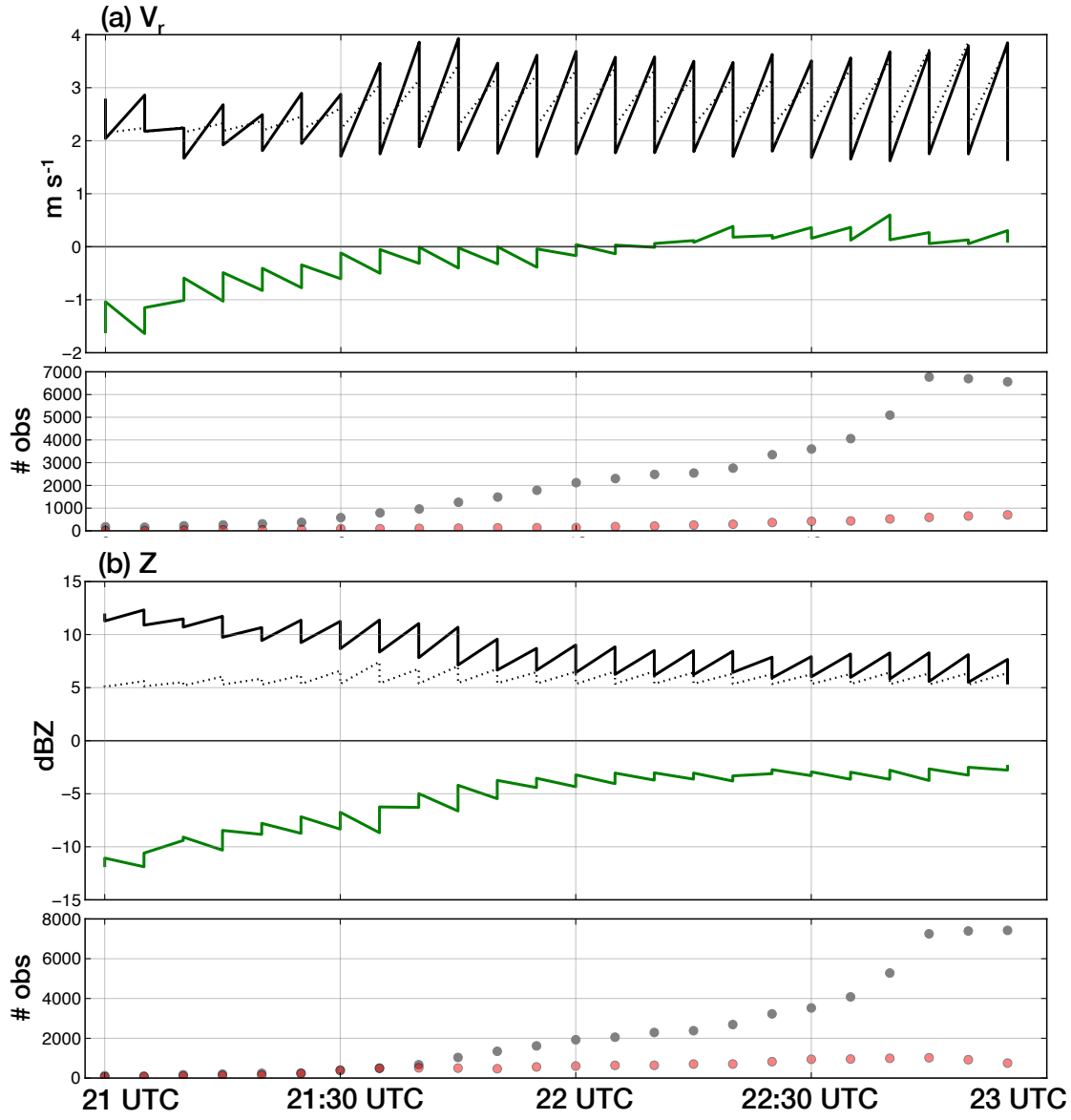

Figure 4.30: Observation-space diagnostics for (a) radial velocity and (b) reflectivity observations in SFCRAD5H. Only those observations that were assimilated are used to compute the diagnostics. RMSI (solid black), mean innovation (solid green), and total spread (dotted black) are plotted in the top part of (a) and (b). The prior and posterior values are plotted together, resulting in the sawtooth pattern. The number of observations assimilated (gray circles) and rejected (red circles) at each assimilation time are plotted in the bottom half of (a) and (b).

velocity is likely due to the larger prescribed value of observation error for reflectivity (5 dBZ error variance). Both the prior and posterior total spread magnitudes are smaller than RMSI during the entire assimilation period, although this is improved later in the period. Reflectivity biases are initially large and negative as convection is spun-up in the ensemble analyses, but stabilize near -3 dBZ for the second half of the assimilation period. This persistent negative bias between 22 and 23 UTC will be investigated in section 4.6.2.

4.6.1.2 Vertically-averaged statistics

To assess variations of the statistics in the vertical, reflectivity and radial velocity observations were aggregated into 500-m vertical bins. Vertical RMSI, spread, and bias profiles for the final prior analysis time (23 UTC) illustrate some of these variations (Fig. 4.31). Radial velocity RMSI is smallest just above the surface between 1 and 4 km (approximately 2.5 m s^{-1}; Fig. 4.31a). Above 4 km, RMSI increases to approximately 3 m s^{-1}. Errors increase rapidly above 10 km due to the small number of observations and a reduced ability to observe the radial wind components. The total spread is larger than the RMSI at all vertical bins below 4 km, while there is a better correspondence between the total spread and RMSI between 4 km and 10 km. A small positive bias exists below 6 km, reaching a maximum near the surface with a local maximum between 3 and 4 km. A negative bias is present aloft between 11 km and 12 km.

The largest reflectivity RMSI occurs at upper-levels (above 10 km where RMSI increases to >10 dBZ) and near the surface (RMSI approximately 8 dbZ; Fig. 4.31b). A minimum is reached in mid-levels between 4 and 6 km. A relative RMSI maxima is also present near 3 km. Total spread is in better agreement with RMSI below 6 km, although discrepancies do exist, most notably near the surface and between 2 and 3 km. Total spread is consistently smaller than RMSI above 6 km, with differences

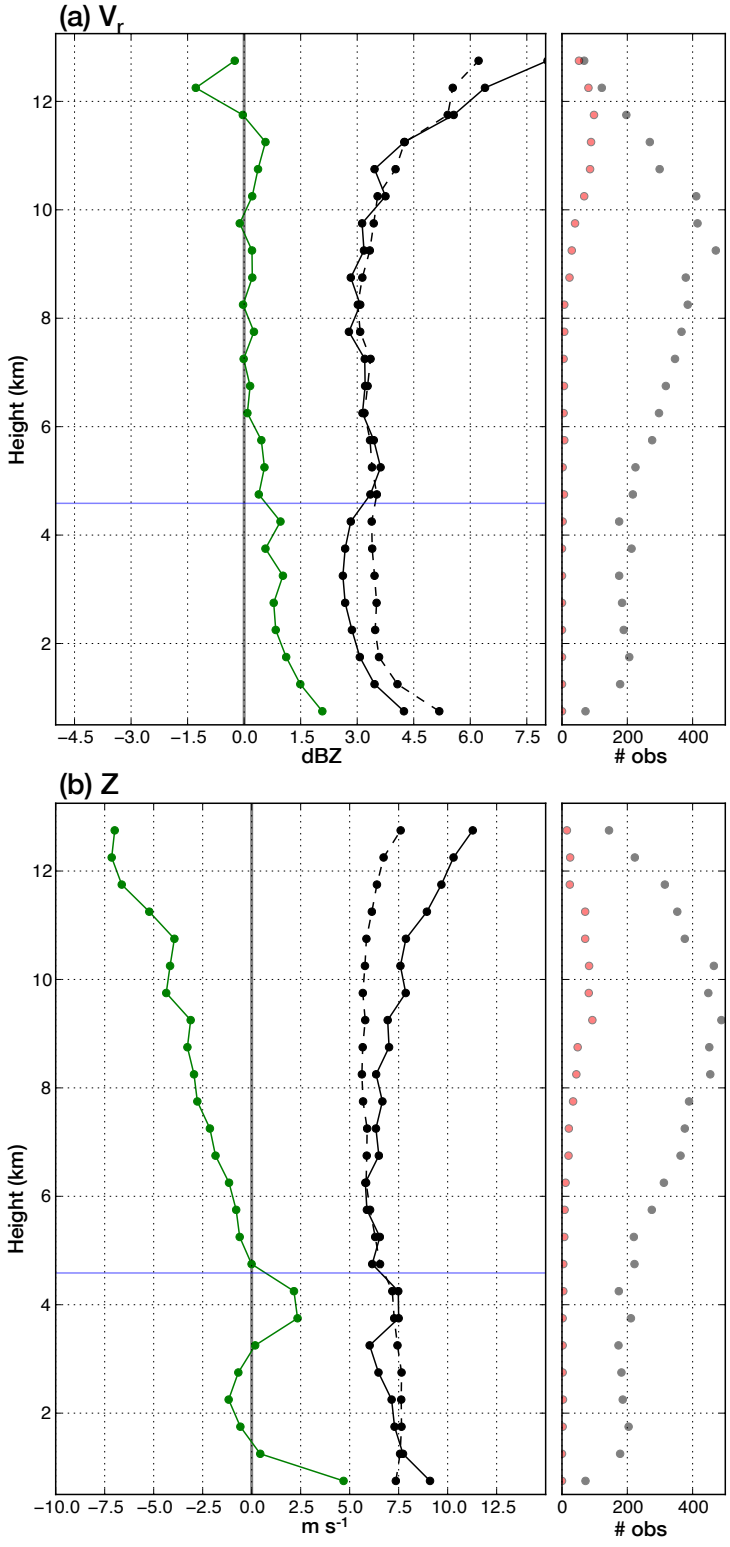

Figure 4.31: As in Fig 4.30, but only for the prior analyses, with statistics aggregated into 500-m height bins. The environmental melting level from the 00 UTC 30 May 2012 KOUN sounding is denoted by the blue horizontal line.

between the two growing with height. These differences in RMSI are strongly affected by the bias. Only the part of the profile between 1 and 3 km is near zero. A +5 dBZ bias is present in the lowest bin of reflectivity observations, and a +2.5 dBZ bias exists in the layer between 3 and 4 km. Above 4 km the bias becomes negatives and its magnitude increases with height. Reflectivity biases above 10 km increase to nearly 8 dBZ. These reflectivity biases are the primary reason RMSI also increases above 6 km. Understanding these systematic biases in the vertical for radial velocity and reflectivity will be the focus of the following section.

4.6.2 Sources of analysis bias

Prior state observation-space biases can exist for various reasons including instrument miscalibration, model error, including errors in model physics parameterizations, and errors in the forward operator connecting the model state to the observations. Since the DA update assumes both unbiased observations and forecasts, conducting DA with these biases present results in a suboptimal data assimilation update and leads to posterior analyses that possess the biases of the forecasts and observations. Thus, interpreting these biases should provide a deeper understanding of analysis and forecast quality.

4.6.2.1 Radial velocity biases

The vertical profiles of bias described in the previous section are plotted for each prior analysis time to gauge temporal trends in bias through the assimilation period (Fig. 4.32; the column at 23 UTC corresponds to Fig. 4.31). The first hour of DA (21 UTC to 22 UTC) will be disregarded since the statistics from the previous section indicated the DA system takes approximately this length of time to stabilize. Beyond this spin-up period, V_r biases are near zero through most of the vertical column except in two areas: above 11 km MSL and near the surface between 0.5 and

3 km MSL. In the former case, a negative V_r bias exists, due to a small number of V_r observations and a limited ability to retrieve the wind field at upper-levels. The latter V_r bias develops after 22:15 UTC and is positive, continuing through the end of the DA period, although reducing in magnitude after 22:45 UTC. While the overall distribution of V_r innovations is approximately Gaussian (Fig. 4.33a), the positive V_r bias at low-levels results in a distribution mean near 1 m s^{-1}.

The positive V_r bias is largest in the 1-2 km MSL layer with a magnitude between 2 – 2.5 ms^{-1}. The V_r observations with the largest positive innovations within this layer are associated with the convection that developed in CI2 between 22:25 UTC and 22:55 UTC (Fig. 4.34). Many V_r innovations are $>$ 8 ms^{-1} in this region, leading to the positive bias in Fig. 4.32a. These innovations are located within and between two storms that merge during the final 30-minutes of the DA period. The motion associated with the northern storm (storm N in Fig. 4.34) is toward the SE, while the southern storm (storm S in Fig. 4.34) motion is toward the NE; innovations greater than 8 m s^{-1} are present along the northeastern edge of storm S around 22:45 UTC (Fig. 4.34c), suggesting that the model is unable to capture the low-level wind field associated with this storm merger.

Most of the innovations >8 m s^{-1} occur when the radial velocity is negative (i.e. away from the radar; Fig. 4.35), which is consistent with the storm motion away from the nearest radar (KVNX) during the storm merger. The prior mean forecast values associated with the positive innovations >8 m s^{-1} fall into two regimes. First, prior forecast values that had the same sign as the observations (negative), but had a smaller magnitude. That is, the prior forecast radial velocity magnitude was smaller than observed, but was in the correct direction. A majority of the largest innovations are due to this type of error. A smaller batch of innovations occured when the prior forecast radial velocity was in the opposite direction compared to the observation, while the magnitude is similar. In the first case, with cloud base near 1.5 km MSL,

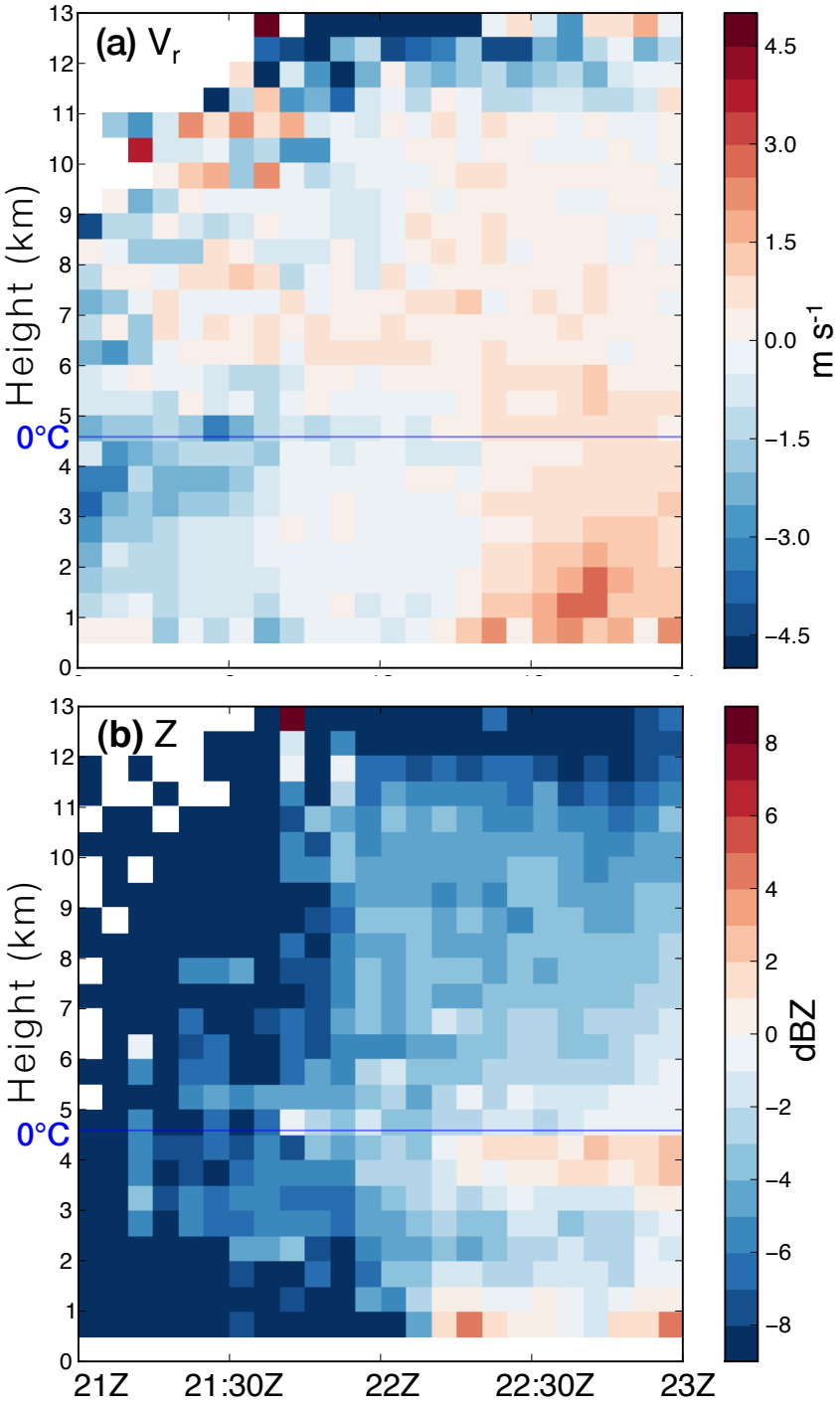

Figure 4.32: Ensemble mean prior bias for (a) radial velocity and (b) reflectivity aggregated by height into 500-m bins for each assimilation time. The environmental melting level from the 00 UTC 30 May 2012 KOUN sounding is denoted by the blue horizontal line.

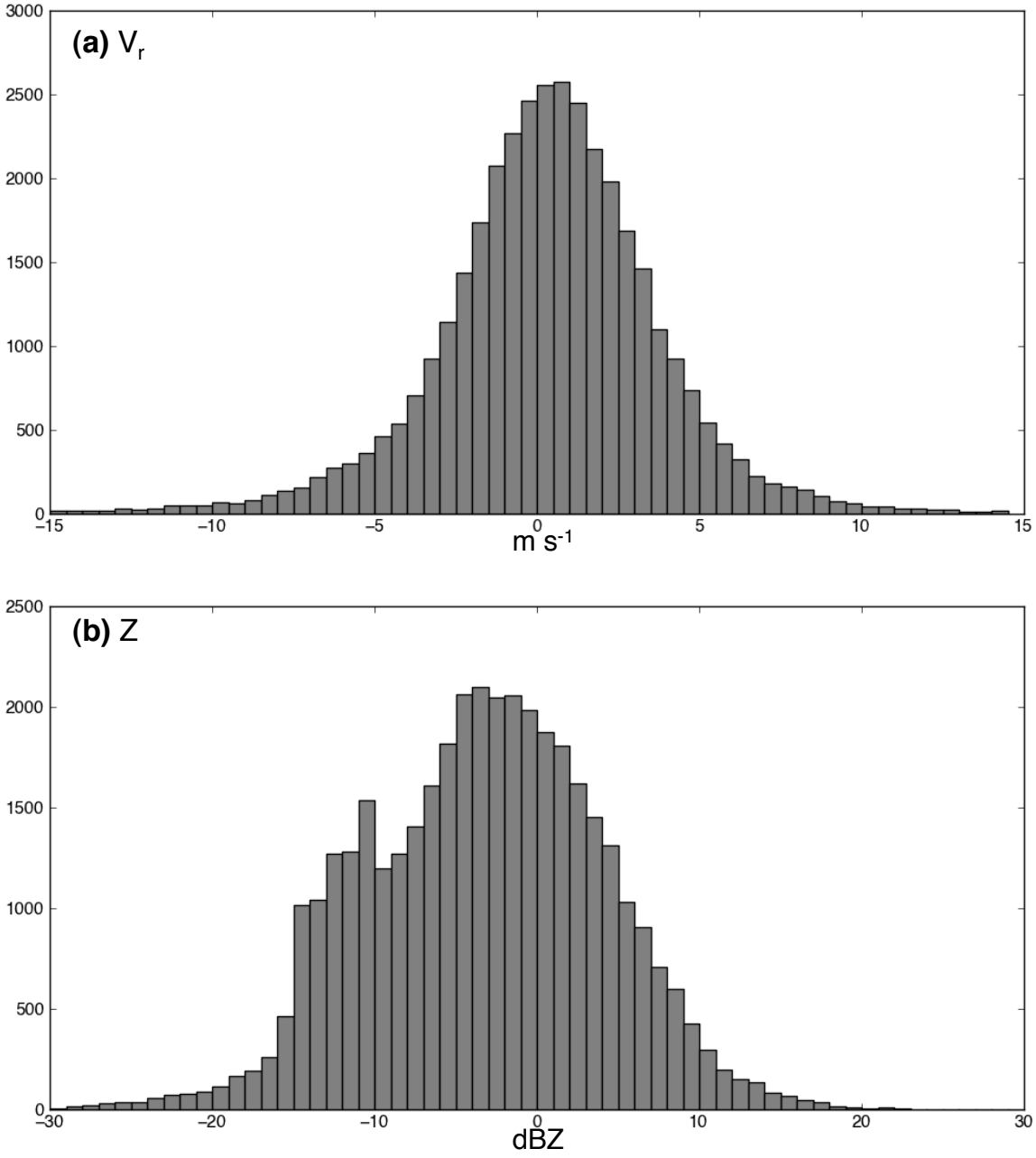

Figure 4.33: Histogram of innovations for (a) radial velocity and (b) reflectivity for observations between 22:30 UTC and 23:00 UTC.

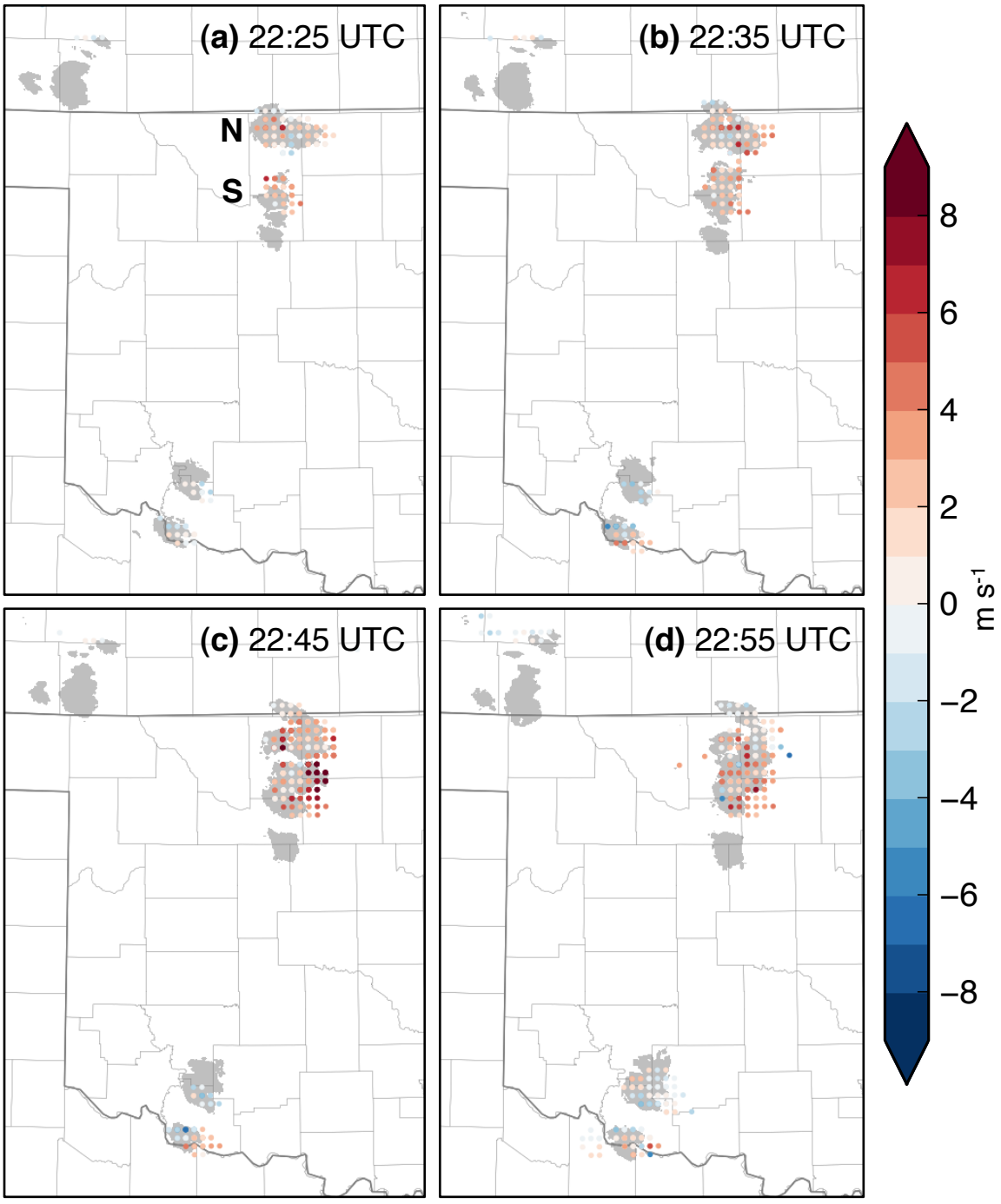

Figure 4.34: Radial velocity observations between 1 km and 2 km MSL at (a) 22:25 UTC, (b) 22:35 UTC, (c) 22:45 UTC, and (d) 22:55 UTC color-coded by the prior innovation associated with each observation. CREF >40 dBZ is shaded.

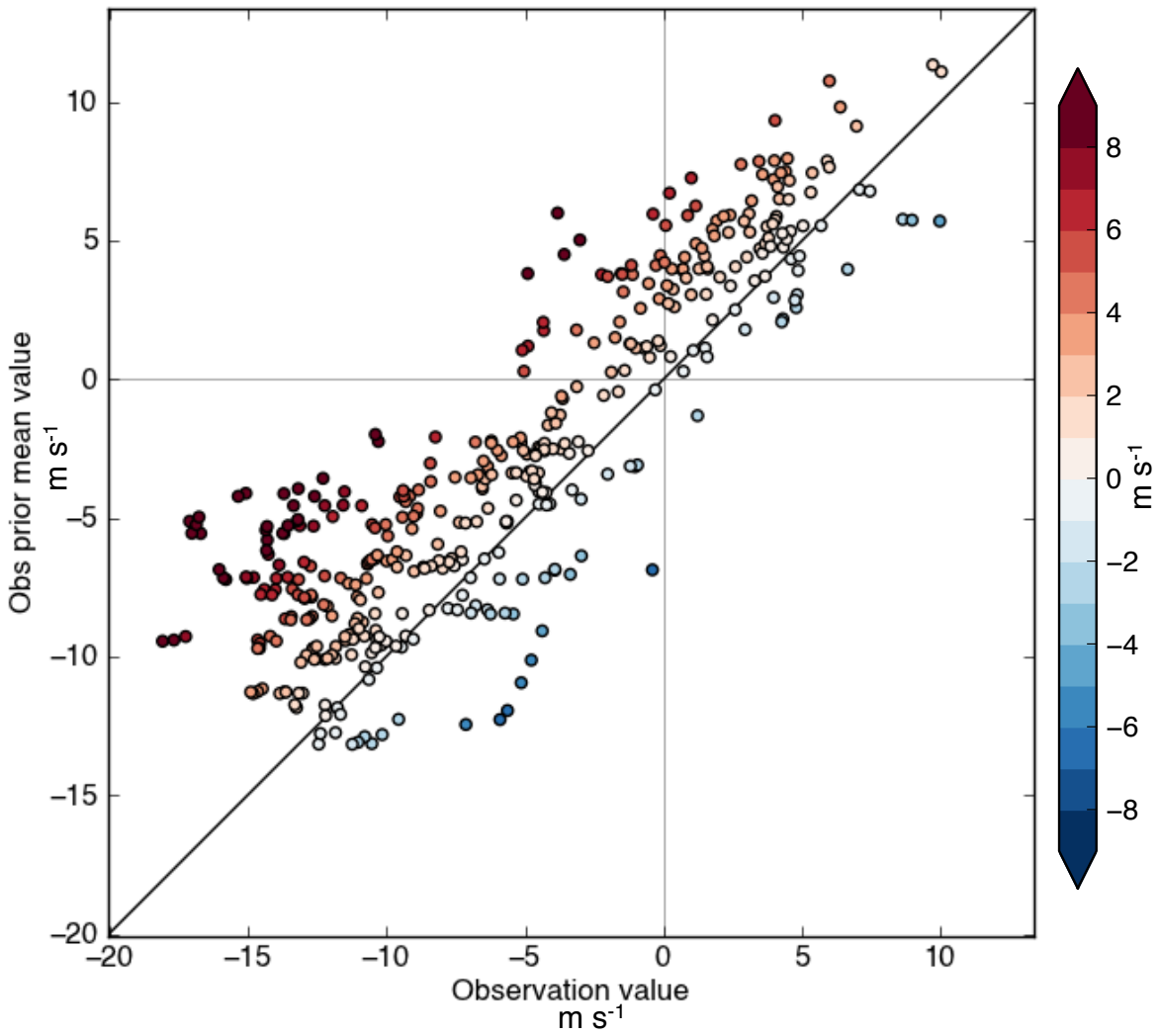

Figure 4.35: Scatter plot of radial velocity observations at 22:45 UTC between 1 km and 2 km MSL and their associated prior mean value (same observations as shown in Fig. 4.34c). Each point is color-coded by the innovation associated with each observation (prior mean value - observed value).

the model likely underpredicted the magnitude of the wind field in the outflow of storm N or S, while in the second case the direction was poorly predicted, potentially due to a displacement error in the position of the surface cold pools associated with these merging storms. Given the complex nature of the convective evolution within CI2, and the transient nature of the bias (the positive radial velocity bias in this layer decreases to near zero by 23 UTC), it is hypothesized that the positive bias is mainly due to predictability limitations, with the true evolution of the cell merger falling outside the envelope of solutions within the ensemble, resulting in a positive radial velocity bias for a brief period of time (similar in nature to the large mean innovations present at the beginning of the DA period due to model spin-up).

4.6.2.2 Reflectivity biases

The vertical structure of Z bias is consistent between 22 UTC and 23 UTC (thus the vertical bias profile at 23 UTC presented earlier in this section is characteristic of the profile through this hour-long period). Positive mean Z innovations occur near the surface and between 3 and 4 km MSL, while above 6 km MSL, the mean Z innovations is negative, with increasing magnitude with height. Mean innovations in each of these regions will be discussed in turn below.

Surface Z bias A positive Z bias is present near the surface (below 1 km AGL) starting around 22:15 UTC. The magnitude of this bias varies through the rest of the DA period, but at several points in the DA period approaches +4 dBZ. This surface Z bias is largest for Z observations between 10 – 20 dBZ, where the magnitude is >than 8 dBZ by 23 UTC (Fig. 4.36a). The positive Z bias extends through a deeper layer for smaller Z observations (Fig. 4.36a), and shallower layers larger Z observations (Fig. 4.36b-d).

A plausible explanation for this bias is due to error from the double-moment microphysics scheme used in the forecast model. Double-moment schemes allow

for hydrometeor sedimentation (the tendency for the mean hydrometeor size to increase toward the ground due to terminal velocity differences). This size-sorting process, while usually a transient phenomenon, can be sustained in environments with wind shear (Kumjian and Ryzhkov 2012). Milbrandt and Yau (2005) described the tendency for double-moment microphysics schemes to overestimate mean rain drop-sizes near the surface due to the size sorting process (see their Fig. 3t). This leads to a positive bias in reflectivity (their Fig. 3n). To counter runaway growth of drop sizes near the surface, some double-moment schemes have placed an upper limit on mean drop diameter, or incorporated a drop-breakup parameterization that acts on drops larger than a prescribed threshold. While the microphysics scheme used herein employs the latter process, uncertainties exist in the formulation and choice of breakup threshold (Morrison et al. 2012). Thus it seems reasonable to suggest a small positive bias of reflectivity near the surface is due to size-sorting. Radar echoes produced by anomalous propagation (i.e. ground clutter) could also produce a bias near the surface, but is likely not a culprit for the surface reflectivity bias since clutter would produce a negative bias (i.e. observed reflectivity larger than forecast reflectivity).

Mid-level Z bias The positive Z bias between 3 and 4 km occurs immediately below the environmental melting layer (Fig. 4.32b), which is approximately located at 4.3 km AGL (determined from 00 UTC 30 May OUN radiosonde). Similar to the surface Z bias, this bias develops after 22:15 UTC and ranges between +1 and +3 dBZ during the second hour of DA. The largest bias occurs with Z observations between 10 and 30 dBZ (Fig. 4.36b-d), since the reflectivity within the bright band typically falls within this range, although it is also present to some degree for all reflectivity observations between 10 and 40 dB (Fig. 4.36a-d). Given the large negative bias for most Z observations $>$ 40 dBZ, this bias near the melting layer is negative, but is 4

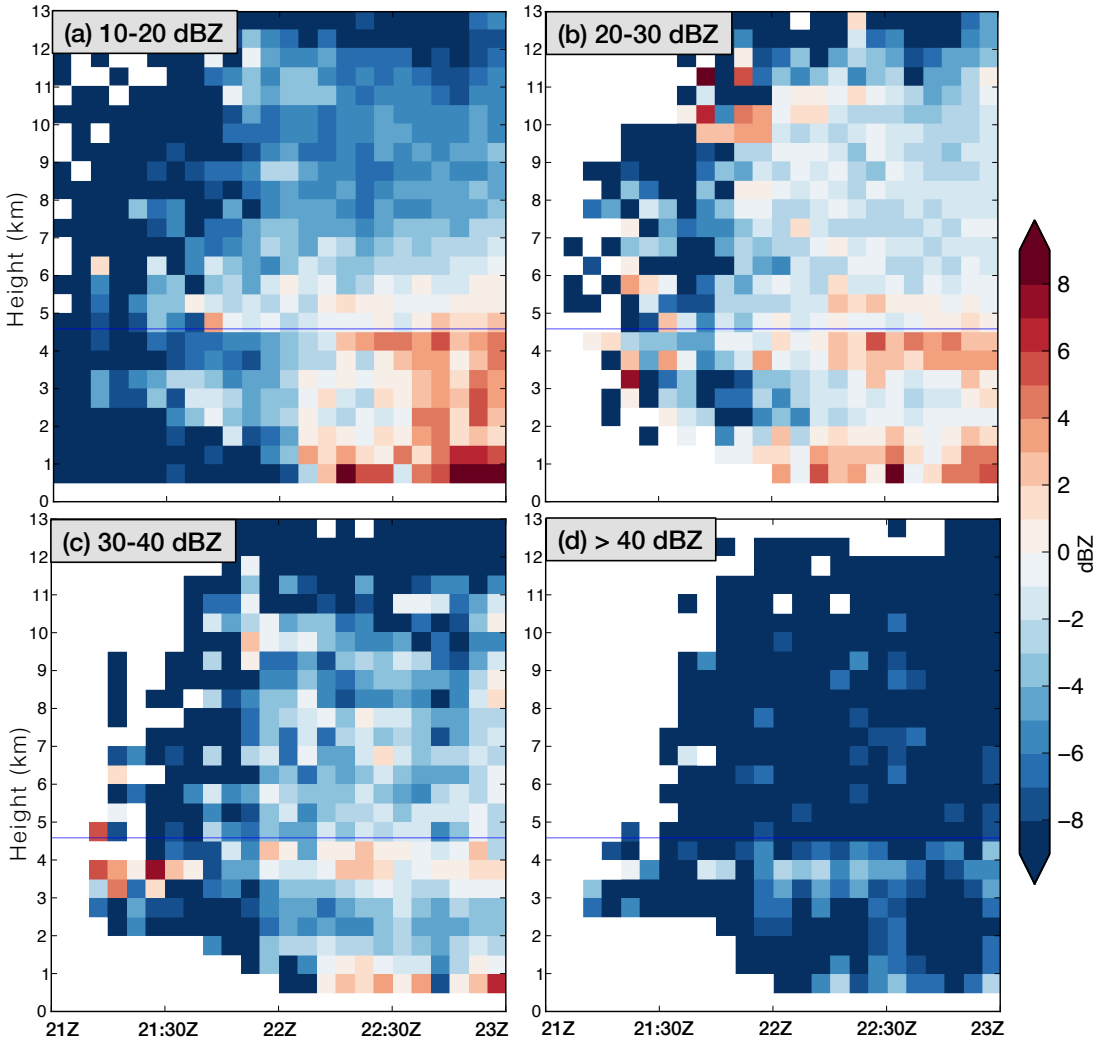

Figure 4.36: As in 4.32, but for reflectivity observations between (a) 10 – 20 dBZ, (b) 20 – 30 dBZ, (c) 30 – 40 dBZ, and (d) >40 dBZ.

– 6 dBZ smaller in magnitude (multiple sources of bias are present that reduce the bias for these observations).

Given its proximity to the melting layer, this positive bias is likely related to the process of melting ice (both snow and graupel) that occurs in the layer below the melting level. While the microphysics scheme incorporates the physical processes of melting and the transition of mass between species that results from the melting process, it does not provide information pertinent to the computation of reflectivity within the melting layer, namely the amount of meltwater that is retained on the

particle exterior during melting, compared to the amount that is absorbed into the interior of the particle. The effect of meltwater on the surface of the ice particle produces the characteristic bright band in radar imagery. Since a break exists between the levels where the largest Z bias exists at mid-levels and near the surface, the surface Z bias is not due to melting graupel that extends to the ground (in addition, vertical profiles of graupel mixing ratio indicate that little to no graupel reaches the ground).

The increase in reflectivity due to ice melting is modeled in the forward operator for reflectivity using the microphysical mixing ratios and temperature profile (see Appendix B for details). Yet, many assumptions must be made when computing reflectivity from melting ice particles, including the fraction of meltwater that is left on the surface and the amount of meltwater that is absorbed into the particle during the melting process. These fairly simple assumptions are associated with significant uncertainty and are likely responsible for the positive bias in reflectivity. This bias gives an indication of the degree of error present in the melting assumptions within the forward operator. One option is to reduce the fraction of meltwater left on the exterior to reduce the predicted reflectivity, or implement a more complex model for melting in the forward operator such as those found in Jung et al. (2010) or Dawson et al. (2013). These tests are left to future work.

Upper-level Z bias A large negative Z bias exists above 6 km (Fig. 4.32b). This bias is dominated by 10 - 20 dBZ reflectivity observations (Fig. 4.36a); most of the observations above 6 km are within this range and are within the convective anvils. For example, the bias above 6 km for 20 - 30 dBZ observations is generally much closer to zero. For higher observed values of reflectivity, fewer observations are present in upper-levels thus bias estimates are often noisy (Fig. 4.36c-d). As was performed for the radial velocity innovations, the locations of observations from 23 UTC were plotted for observations within the 8 - 10 km layer with values of reflectivity between 10 - 20 dBZ (Fig. 4.37). Within this layer, most observations are located within

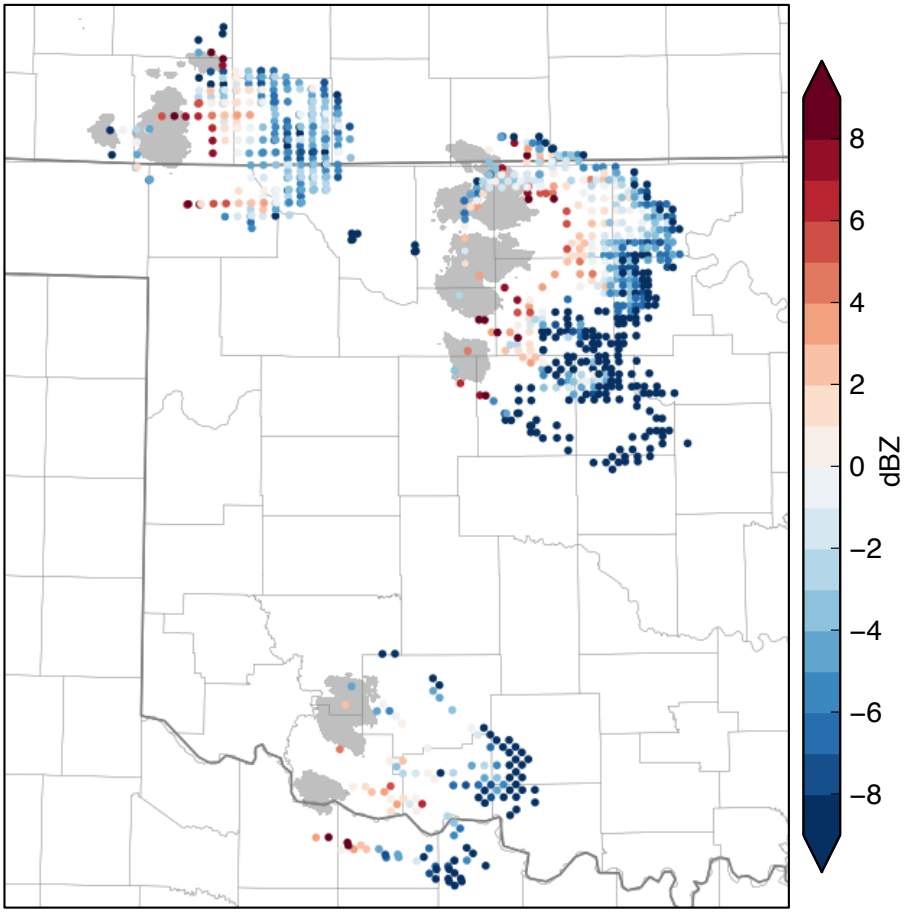

Figure 4.37: As in Fig. 4.34, but for reflectivity observations between 8 km and 10 km MSL at 23:00 UTC.

the anvil region, which spreads toward the ESE away from the convective cores. Specifically, the observations with the largest negative innovations (greater than 8 dBZ) are located along the outer edge of the anvils, and result in a bimodal peak in the distribution of reflectivity innovations (4.33b). Here, the negative bias indicates that the anvils associated with convection in the prior forecasts have not spread as far downstream as the observations indicate.

One hypothesis for the reflectivity bias is a slow bias in the environmental wind profile. Due to this bias, storm motions were slower than observed in the forecasts and, at anvil-level, cloud ice and snow were not advected far enough downstream during the forecast. This is supported by the ensemble hodographs in CNTL and SFC3H valid at 23 UTC 29 May 2012 from the nearest grid point to the OUN sounding

location (Fig. 4.24). The ensemble mean wind speed within the 8 - 10 km layer varies from 18 m s^{-1} to 27 m s^{-1} at 8 km MSL and 10 km MSL, respectively, while the observed wind speed is 22 m s^{-1} at 8 km MSL and 31 m s^{-1} at 10 km MSL. Thus, the ensemble mean wind speed is approximately 4 m s^{-1} less than the observed wind speed within this layer (the differences in the forecast and observed wind direction are less significant).

Assuming these differences between the forecast and observed profiles are similar across the domain, the integrated effects of this forecast error during the two-hour DA period could produce the large negative reflectivity biases along the downstream edge of the convective anvils. After 2 hours of radar DA, the 4 m s^{-1} anvil-level flow bias would lead to an upstream displacement error of approximately 30 km; this is similar to the width of the downstream anvil edge composed of the largest negative innovations.

A second hypothesis involves an underprediction of anvil-level ice and snow due to incorrect particle fall speeds within the microphysics parameterization. That is, as snow is advected downstream by the anvil-level winds, it tends to fall out more quickly than what is observed, underpredicting the snow mixing ratios leading to smaller than observed reflectivity. This is partly supported by the gradual increase in reflectivity bias that occurs in the anvils of the northern convection within CI2. Here, the reflectivity bias is actually positive closer to the convection core and becomes more negative further downstream within the anvil. A definitive explanation for these errors are left to future work, but its impact, and the impacts of the other reflectivity biases on optimal data assimilation, will be discussed in the following section.

4.6.2.3 Discussion

The various biases described in the previous sections can produce undesirable and deleterious effects on the resulting analyses and forecasts. For example, in the present

experiments, the positive bias associated with the reflectivity forward operator within the melting layer results in prior innovations that are predominantly positive. For reflectivity, this implies a positive adjustment to the rain, snow, and/or graupel mixing ratios since reflectivity observations and hydrometeor mixing ratios are positively correlated. Since the bright band develops only once appreciable ice hydrometeors develop, these errors likely play a less significant effect early in the development of convection(Fig. 4.32b). But once the melting process begins to occur within the forecast, assimilation of reflectivity observations, through positive correlations with the microphysical state, will result in an addition of mass to the hydrometeor species, solely due to the biased forward operator. In a cycled assimilation system such as the one used herein, these errors are propagated to the next assimilation time. Thus, a persistent continual increase in hydrometeor mixing ratios will continue to occur at every assimilation time within and surrounding the layer possessing melting ice hydrometeors. A similar behavior can occur with errors due to a biased forecast model.

4.6.3 Ensemble forecasts from SFCRAD5H

While the focus of the previous section was on the fidelity of the SFCRAD5H prior analyses produced during the 21 UTC to 23 UTC DA period, this section will examine the 50-member 3-hour ensemble forecasts initialized with the final 23 UTC posterior analyses. The ensemble forecasts provide an additional means to assess analysis quality, as reduced analysis error is expected to translate into gains in forecast skill.

For verification, a similar approach is taken to that applied with the surface DA experiments. This includes comparing forecast CREF areas with observed CREF by using the grid-point maximum value of CREF computed over a one-hour period or the entire three-hour forecast period. Probabilities of CREF exceeding 25 dBZ between 23 UTC and 02 UTC are generated with these 1-hr and 3-hr maximum fields

(PROB1H-CREF25 and PROB3H-CREF25, respectively). This same procedure was applied to the observed CREF dataset, although for these comparisons an observed CREF threshold of 40 dBZ is applied to isolate the track of each convective core (since convection has matured by 23 UTC, using the observed 25 dBZ CREF contour would include large areas of anvil).

To identify the tracks of the most intense convective cores in SFCRAD5H, the maximum upward vertical velocity (UVV) between the surface and 400 hPA is recorded at each time-step and a time-maximum at each grid point is applied to produce an hourly-maximum UVV field (analogous to the hourly-maximum CREF field). Hourly-maximum UVV is used in addition to hourly-maximum simulated CREF, including both one-hour and three-hour UVV probabilities (PROB1H-UVV10 and PROB3H-UVV10, respectively) and the one-hour and three-hour ensemble maximum UVV (MAX1H-UVV and MAX3H-UVV, respectively). Verifying the probabilistic forecasts is challenging since only one event is analyzed, thus, the comparisons between forecasts and observations will be subjective. Further, analysis will focus on the three areas of most intense convection in the domain (the southeastward moving supercells in north-central OK, the left-split and right-split originating within CI3; Fig. 4.5)

Overall, the SFCRAD5H forecast fields of PROB3H-CREF25 and PROB3H-UVV10 capture the most intense observed convection within the domain (Fig. 4.38). In the northern half of the domain, a NW-SE axis probability maximum is collocated with the track of several supercells that propagated toward the SE between 00 UTC and 02 UTC. These storms developed within the southern portion of CI2 and produced very large hail and a brief tornado, and were among the most intense storms during this event. The orientation and magnitude of the probability axis provides confidence in the corridor where these supercells would eventually track during the forecast period. Within this corridor, convection is forecast to move slower than

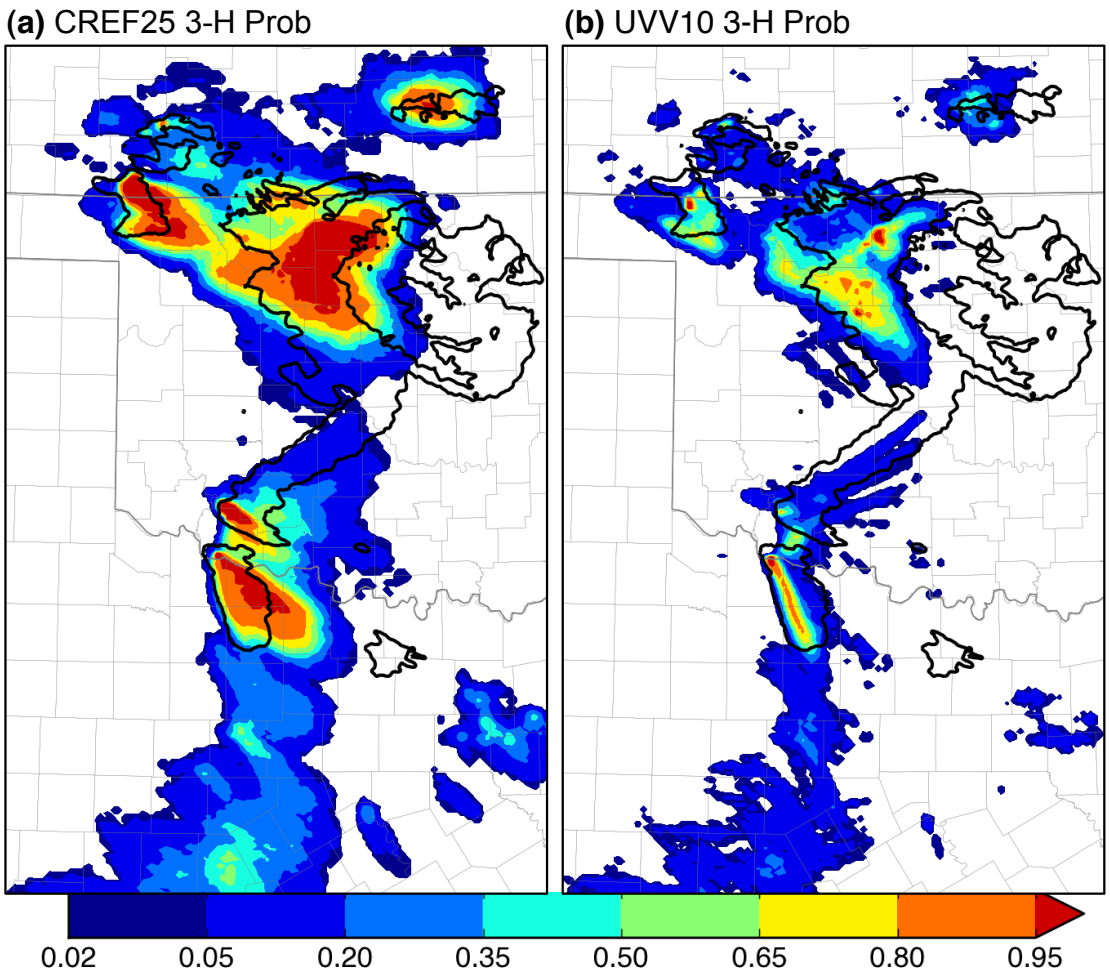

Figure 4.38: 23 UTC to 02 UTC probability of (a) CREF >25 dBZ (PROB3H-CREF25) and (b) UVV >10 m s^{-1} (PROB3H-UVV10). The areas where one-hour maximum observed CREF >40 dBZ is denoted by the solid black contour.

observed, likely due to the previously mentioned environmental wind speed bias (Fig. 4.24). For example, by 02 UTC the observed right-moving supercell was in the process of merging with the left-moving storm originating in CI3, but the PROB3H-CREF25 axis suggests that no members forecast this merger to occur before the end of the forecast period. This bias affects, to some degree, all the convective-scale forecasts in this chapter and illustrates how mesoscale errors can limit convective-scale forecast skill. The PROB3H-UVV10 swath is almost perfectly co-located with the observed CREF25 contour (Fig. 4.38). Timing errors for this storm are less than other areas of convection. Given its predicted intensity, a slower storm motion due to internal storm processes likely led to smaller three-hour forecast errors (Bunkers et al. 2000).

SFCRAD5H also does a poor job forecasting the development of a line of convection in NE OK after 01 UTC. This line grows upscale and is responsible for severe wind gusts across eastern and southeastern OK after 02 UTC, but this portion of the case will not be the focus of any analysis since none of the experiments conducted herein were able to capture its development, although several members do predict its development after 02 UTC. Future work should investigate the ability of radar DA beyond 23 UTC to capture and predict this part of the convective system.

The PROB1H-CREF25 field between 23 UTC and 00 UTC demonstrates the effectiveness of radar DA at the beginning of the forecast period (Fig. 4.39a). Higher values of PROB1H-CREF25 (>95%) are colocated with areas of active convection, indicating radar DA was successful at sustaining convection during the first forecast hour. Two areas exist where convection appears to be poorly initialized. The southern storms within CI2 are associated with lower PROB1H-CREF25 values during the first hour than the convection to its north. Also, in CI3, the left-split associated with the northern storm in CI3 is associated with smaller PROB1H-CREF25 values than the southern storm. Several ensemble members capture the motion and track of the left-split during the 23 UTC – 00 UTC period, although by the end of the

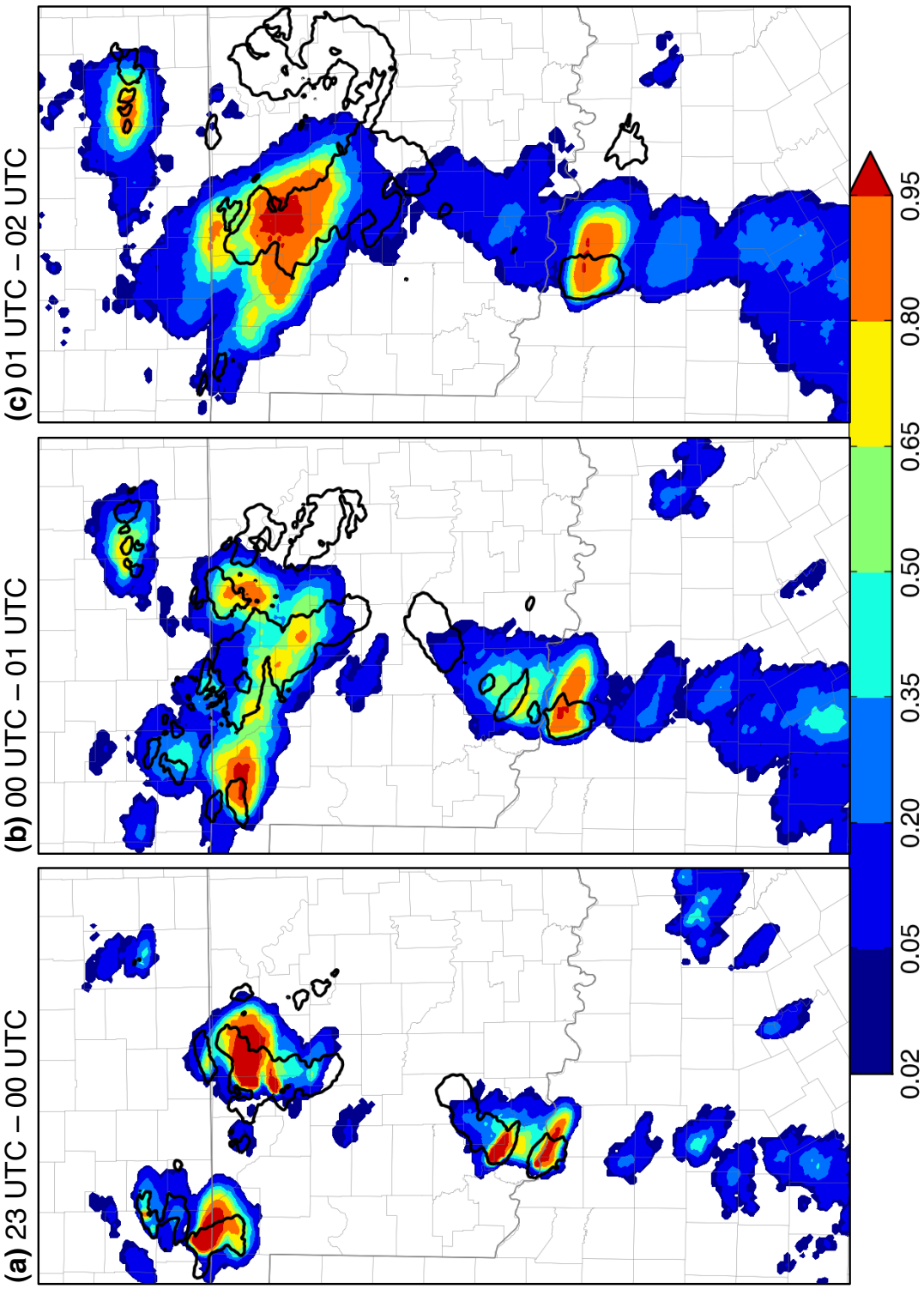

Figure 4.39: As in Fig. 4.38a, but for one-hour probabilities of CREF >25 dBZ (PROB1H-CREF25) and one-hour maximum observed CREF >40 dBZ from (a) 23 UTC to 00 UTC, (b) 00 UTC to 01 UTC, and (c) 01 UTC to 02 UTC.

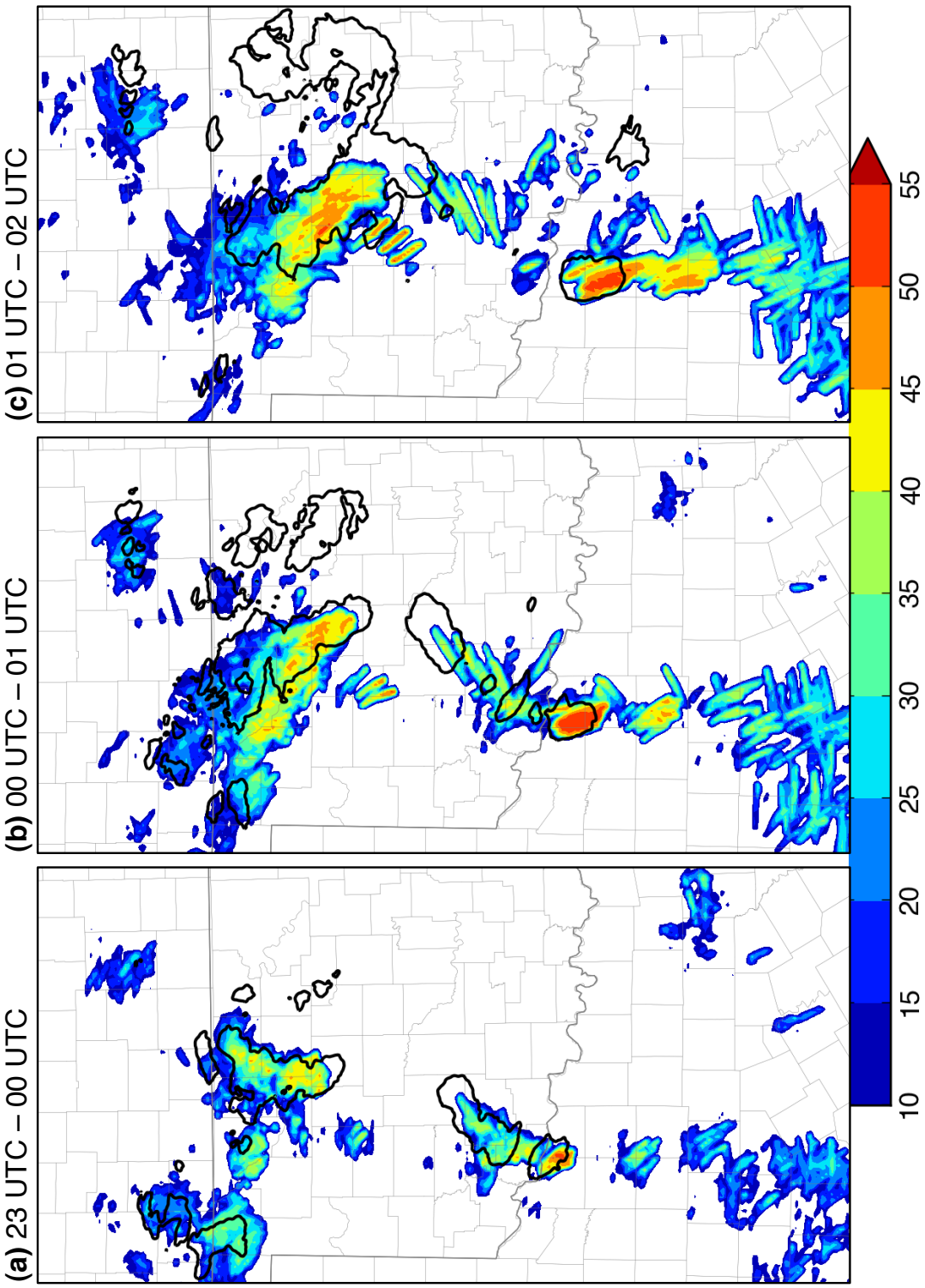

Figure 4.40: One-hour ensemble maximum UVV (MAX1H-UVV; m s^{-1}) and one-hour maximum observed CREF >40 dBZ from (a) 23 UTC to 00 UTC, (b) 00 UTC to 01 UTC, and (c) 01 UTC to 02 UTC.

one-hour forecast, the observed storm location is ahead of the envelope of CREF probabilities. In CI4, PROB1H-CREF25 values are reduced compared to the surface DA experiments (c.f. Fig 4.13b and Fig. 4.38a), due to the assimilation of clear-air observations (non-zero probabilities remain near DFW in a region not affected by clear-air radar observations).

Between 00 UTC and 01 UTC, PROB1H-CREF25 values increase in the area from CI1 to CI2. Here, three probability maxima are present associated with the following: 1) convection originating in CI1 that moves toward the SE, 2) convection in the northern part of CI2 that moves eastward, and 3) new convective development just west of CI2, between CI1 and CI2 (Fig. 4.39b and 4.40b). The southeastern part of the third area is composed of convection that developed in the southern part of CI2, along with new development to its west. This convection is responsible for extending the PROB1H-CREF25 axis further to the southeast compared to the forecasts that only assimilate surface data. SFCRAD5H also captures the convective development in south-central KS. Further south, PROB1H-CREF25 values >90% are co-located with the southern right-moving supercells in CI3, while lower PROB1H-CREF25 values are present in association with the northern storm. SFCRAD5H successfully predicted that the southern storm would intensify after 00 UTC, with MAX1H-UVV values greater than 50 m s^{-1} associated with this convection (Fig. 4.40b). Further, the demise of the northern right-moving storm in CI3, once the southern storm intensified, is well predicted. A third of the ensemble successfully predicts the occurrence of a left-moving supercell, as indicated by PROB1H-CREF25 values >35% extending to the NE toward OKC between 00 UTC and 01 UTC (Fig 4.39b), along with MAX1H-UVV storm tracks (Fig. 4.40b). This left-moving storm is observed to move faster than predicted by the swaths of both PROB1H-CREF25 and MAX1H-UVV.

After 01 UTC, PROB1H-CREF25 values increase within north-central OK (Fig. 4.39c). Again, this convection moves slower than the observed convection

between 01 and 02 UTC, likely due to errors in the cloud-layer environmental wind field discussed previously. The highest PROB1H-CREF25 values (>80%) in CI3 are associated with the right-moving supercell in northern TX. The position of this storm between 01 UTC and 02 UTC is predicted well. In CI3, the PROB1H-CREF25 field extends further downstream than the observed CREF >40 dBZ contour due to the former field capturing the convective anvil. The left-split in CI3 is forecast by approximately a third of the ensemble members to continue moving northeastward through 02 UTC (Fig. 4.39c).

4.6.4 Ensemble forecasts from 22 UTC, 22:30 UTC, 23 UTC

Ensemble forecasts were initialized at 22 UTC, 22:30 UTC, and 23 UTC after 1, 1.5, and 2 hours of radar DA (in addition to surface DA), respectively. As discussed in section 4.5.5, differences between these experiments are due to a combination of the longer DA period and the shorter-forecast lead-time. Yet, the effects of radar DA can be seen through the systematic reorientation of forecast probabilities as radar data are assimilated for longer periods.

As the DA period is extended, PROB3H-UVV10 values increase along the axis of the most intense observed convection across central OK (Fig. 4.41). While this axis extends further toward the SE in the 23 UTC forecast, this axis of highest PROB3H-UVV10 values is slightly biased toward the NW, with convection not reaching the OKC metropolitan area until after 02 UTC (in reality this occurred closer to 01 UTC). In areas surrounding the observed CREF >25 dBZ contour, PROB3H-UVV10 values are reduced as additional clear-air observations are assimilated. Since CI occurred between 21:15 UTC and 21:30 UTC, the forecast initialized at 22 UTC has only been impacted by 30-45 minutes of radar DA (not including clear-air observations), during the early evolution of the storms.

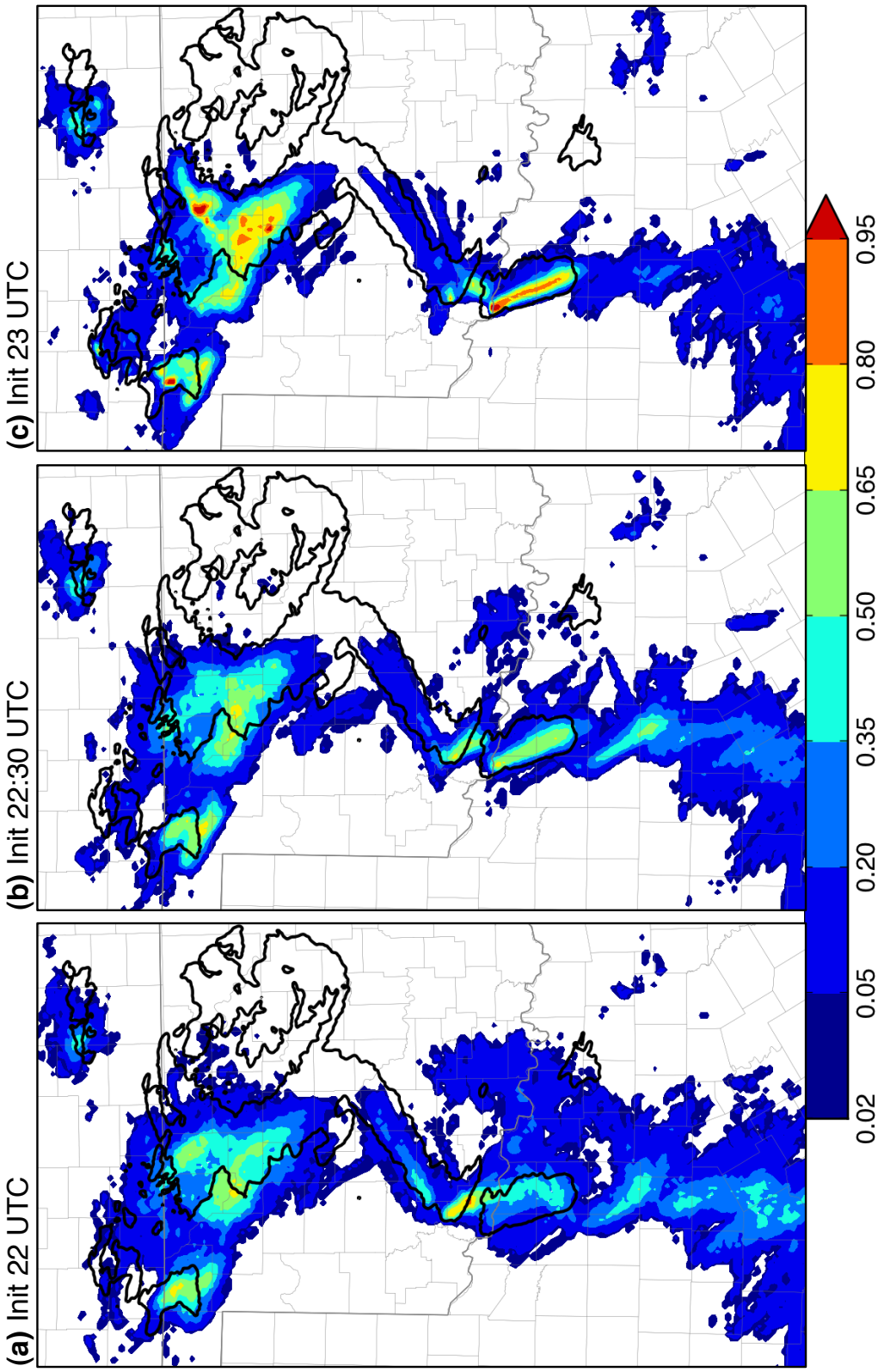

Figure 4.41: As in Fig. 4.38b, but for ensemble forecasts initialized at (a) 22 UTC, (b) 22:30 UTC, and (c) 23 UTC.

Differences between the three forecasts are largest in CI3. In the forecast initialized at 22 UTC, the PROB3H-UVV10 swaths indicate the development of a left- and right-moving pair of storms originating near the position of the observed northern supercell (Fig. 4.41a). Relatively low PROB3H-UVV10 values (<35%) exist along the path of the observed southern supercell. In the 22:30 UTC forecast, PROB3H-UVV10 values increase along the observed track of the southern supercell (Fig. 4.41b), and by 23 UTC, this axis narrows and probabilities increase to >80% (Fig. 4.41c). Further, given the agreement among the ensemble members in the forecast track of the southern supercell in CI3 by the 23 UTC forecast, nearby areas of small PROB3H-UVV10 values that are present in the 22 UTC and 22:30 UTC forecasts are reduced in the 23 UTC forecast, especially in areas west of the track, with PROB3H-UVV10 values decreasing to <20%, in the region near CI4.

The forecast of the observed left-split that emerges from CI3 is sensitive to the initialization time. In the 22 UTC forecast, the northern supercell was long-lived and the southern supercell was not predicted as confidently (Fig. 4.41a), resulting in more members that sustain a left-split originating from the northern supercell. In the 23 UTC forecast, the southern supercell was predicted to be long-lived, while less than half of the ensemble members successfully forecast the development of the northern supercell (Fig. 4.41c). This change in forecast behavior leads to fewer left-splits in the ensemble along the axis where the left-split was observed in the forecast initialized later, even after two hours of radar DA. The differences between the ensemble forecasts suggests that DA has difficulty in developing both the northern and southern supercell in CI3 as was observed, leading to subsequent differences in the predictions of the long-lived left-split. Due to the juxtaposition of these two storms, limitations from the relatively coarse model and observation resolution in these experiments may hamper the ability to accurately predict both storms.

4.6.5 Impact of surface and radar data on 0-3hr forecasts

Two additional experiments were conducted to isolate the effects of the radar and surface DA. In the first, SFC3H is extended by assimilating surface data every 5-minutes for an extra two hours between 21 UTC and 23 UTC (SFC5H). In the second, surface data are withheld; radar DA begins at 21 UTC and continues to 23 UTC (RAD2H), thus the initial 18 UTC ensemble is advanced freely to 21 UTC before radar DA, without any surface DA. SFCRAD5H, SFC5H, RAD2H, and CNTL will be compared using the PROB3H-UVV10 field. Differences between SFC5H (Fig. 4.42b) and SFCRAD5H (Fig. 4.42d) are primarily due to radar DA, while differences between RAD2H (Fig. 4.42c) and SFCRAD5H (Fig. 4.42d) are primarily due to surface DA. To map the response process in space-time, we ran queries on our coded transcripts for instances when a portion of text received codes for both a component of the response process (e.g., awareness, confirmation, protective actions) and any spatio-temporal benchmark.

4.6.5.1 Northern half of domain

When only surface data are assimilated, the PROB3H-UVV10 axis in central OK is centered to the NW of the observed axis of convection (Fig. 4.42b) with the most intense convection straddling the observed CREF boundary (Fig. 4.43b). The maximum of PROB3H-UVV10 near CI1 matches well with the observed CREF area. When only radar data are assimilated (RAD2H), the axis of convection in central OK is centered further toward the SE (Fig. 4.42c and Fig. 4.43c) compared to SFC5H. In addition, a bulls-eye of higher PROB3H-UVV10 values is present to the NE of the main PROB3H-UVV10 axis that is not present in SFC5H. Both of these improvements over SFC5H are due to the assimilation of radar observations associated with CI2. On the other hand, errors exist in RAD2H that do not exist in SFC5H. Convection in CI1 is forecast in RAD2H to persist well into the forecast period, long after this

convection was observed to dissipate (Fig. 4.42c and Fig. 4.43c). Also, the area of low-end PROB3H-UVV10 values (<20%) that fall outside the envelope of observed convection is larger in RAD2H, while the low-end PROB3H-UVV10 maximum in southern KS (near Wichita) associated with observed convection late in the forecast period is better placed in SFC5H. Thus, surface DA imparts a significant contribution to reducing these low-end probabilities, even more so than through the assimilation of clear-air observations.

While radar DA alone does lead to forecast improvements, the RAD2H ensemble retains several detrimental environmental errors present within CNTL, thus RAD2H is handicapped without the addition of surface observations to correct errors in the surface environment, leading to increased convective forecast error compared to SFCRAD5H. The best forecast comes when both datasets are assimilated together (SFCRAD5H). Further, among the three experiments, the spatial pattern of PROB3H-UVV10 and MAX1H-UVV values in RAD2H is most similar to CNTL (c.f. Fig. 4.42a,c), while the pattern in SFC5H is most similar to SFCRAD5H (c.f. Fig. 4.42b,d). In other words, assimilating surface data alone provides gains in forecast accuracy (as primarily assessed by PROB3H-UVV10) that are at least as large, if not larger, than those gained by assimilating radar data.

4.6.5.2 Southern half of domain

The forecast of convection in the southern part of the domain is quite sensitive to the inclusion of the radar and surface datasets. Assimilating surface data alone, as in SFC5H, produces a northern supercell that is long-lived, leading to low, diffuse probabilities along the observed track of the southern supercell where only a few members produce convection (Fig. 4.44b). With radar data alone, as in RAD2H, the southern storm is predicted to be dominant, leading to fewer members that produce long-lived convection where the northern storm was observed (Fig. 4.44c). The

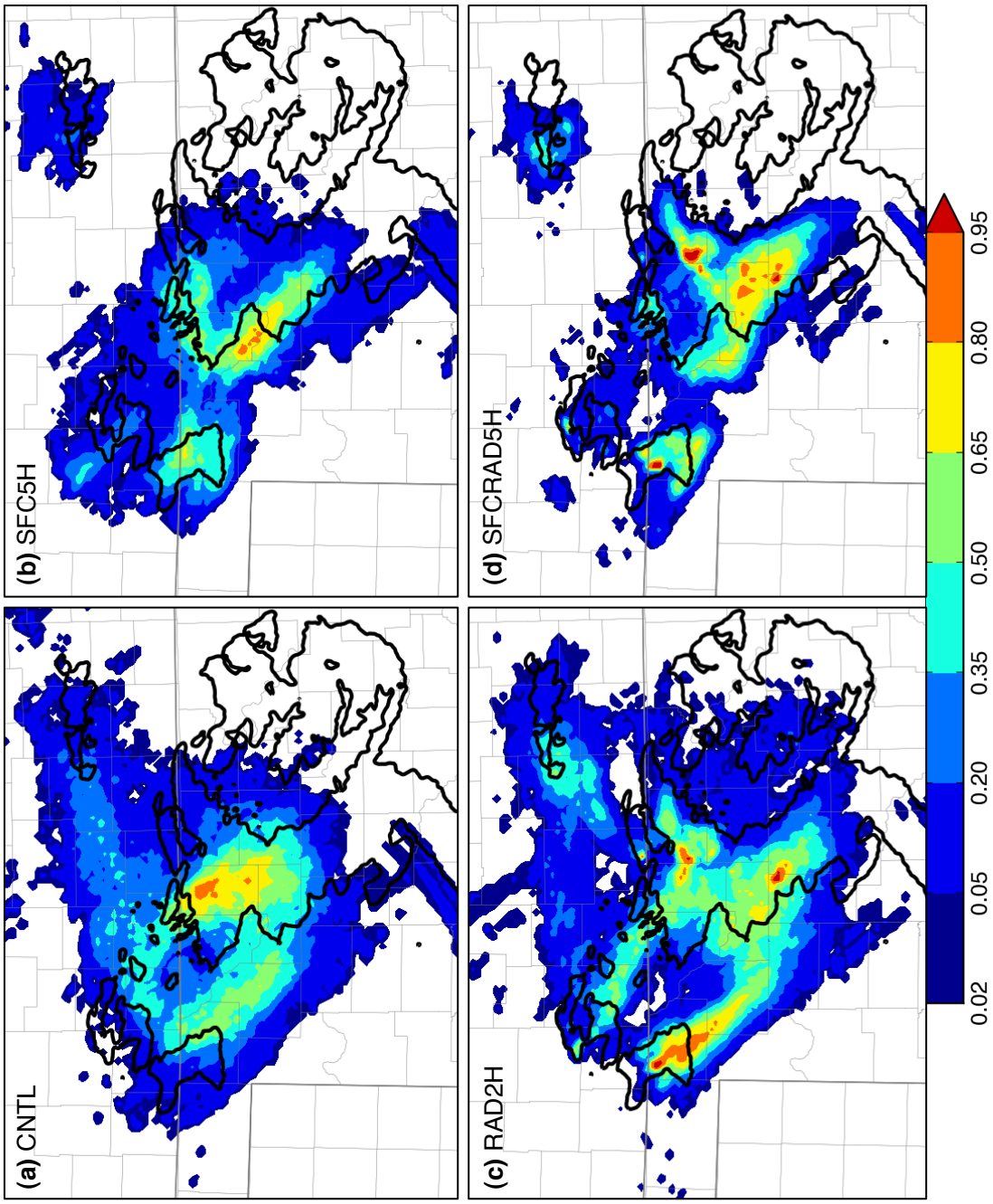

Figure 4.42: As in Fig. 4.38b, but for PROB3H-UVV10 values in the northern half of the domain from (a) CNTL, (b) SFC5H, (c) RAD2H, and (d) SFCRAD5H.

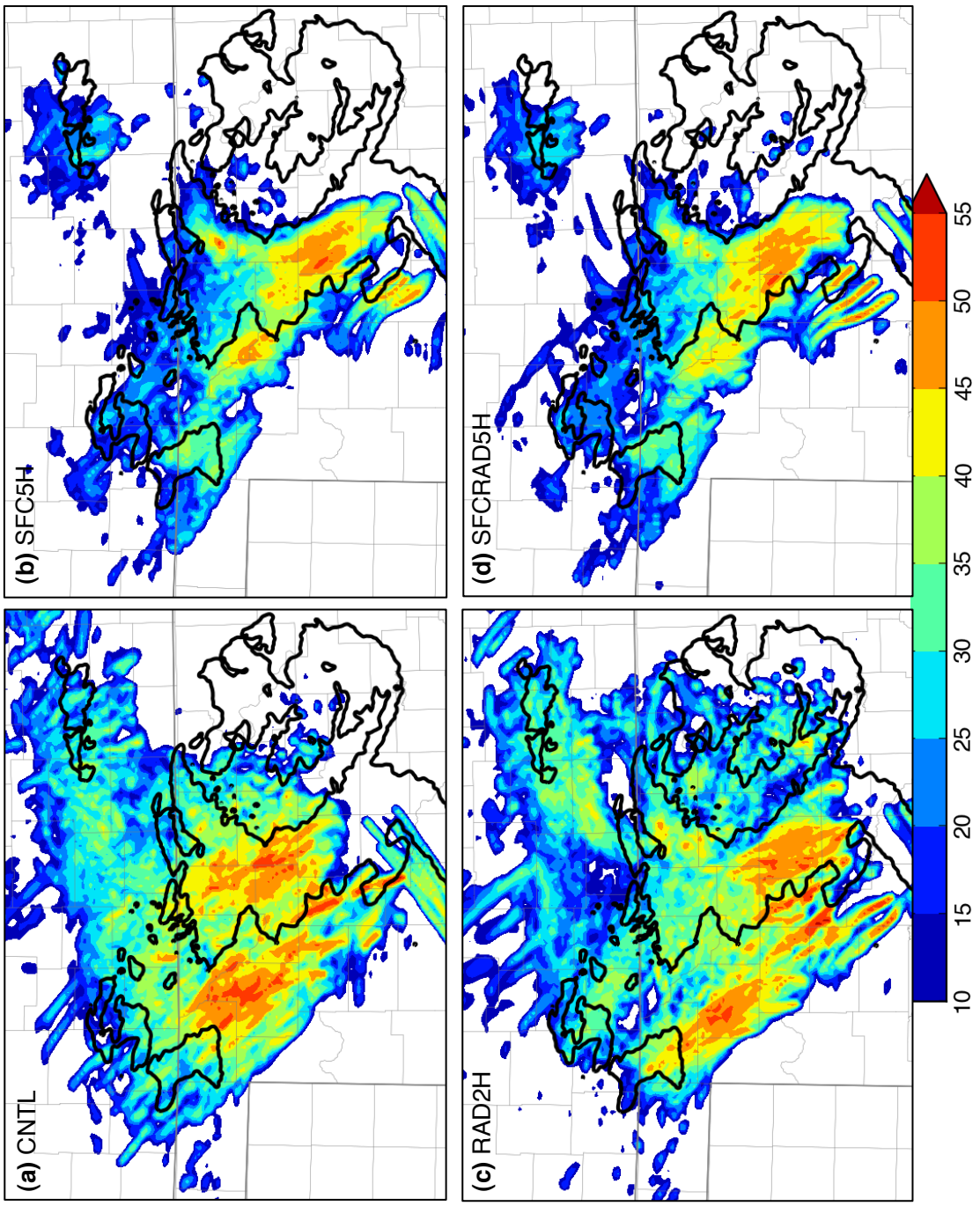

Figure 4.43: As in Fig. 4.40, but for one-hour ensemble maximum UVV >10 m s^{-1} (MAX1H-UVV) in the northern half of the domain from (a) CNTL, (b) SFC5H, (c) RAD2H, and (d) SFCRAD5H. Black contour is three-hour maximum observed CREF >40 dBZ.

combination of surface and radar data in SFCRAD5H produces a narrow swath of large PROB3H-UVV10 values (>80%) in excellent agreement with the observed track of the southern supercell (Fig. 4.44d). The increase in PROB3H-UVV10 values is due to decreased spread in the forecast tracks of convection among the ensemble members (c.f. Fig. 4.45c,d). PROB3H-UVV10 values for the northern supercell are similar in SFCRAD5H and RAD2H. The left-split storm is more favored in SFC5H than RAD2H or SFCRAD5H for similar reasons as discussed in section 4.6.4. In SFC5H, more members agree in the development of a dominant northern supercell, resulting in more tightly clustered storm tracks compared to the RAD2H and SFCRAD2H, increasing PROB3H-UVV10 values (c.f. Fig. 4.45b,d) Finally, in the experiments that assimilated clear-air observations (RAD2H and SFCRAD5H), PROB3H-UVV10 values are reduced where convection was not observed, while in SFC5H, probabilities >35% remain in CI4, similar to CNTL.

These results suggest that the biggest benefit to forecast skill is radar DA in the southern part of the domain, although the combination of surface and radar DA again produces the most skillful forecast in both parts of the domain. Surface DA plays an important role in improving short-term forecast skill in both portions of the domain, with a larger impact in the northern portion of the domain where errors in the prediction of surface fields and placement of boundaries are larger and result in larger errors in convective development between 00 UTC and 01 UTC.

4.6.6 Impact of reflectivity assimilation on 0-3hr forecasts

Given the reflectivity biases described in a previous section, and its non-linear relation to the state, it is possible that the assimilation of reflectivity observations is detrimental to the state estimate. To test the sensitivity of the forecasts to the assimilation of reflectivity, reflectivity observations in precipitation were withheld from the assimilation and a 3-hour 50-member ensemble forecast was initialized from

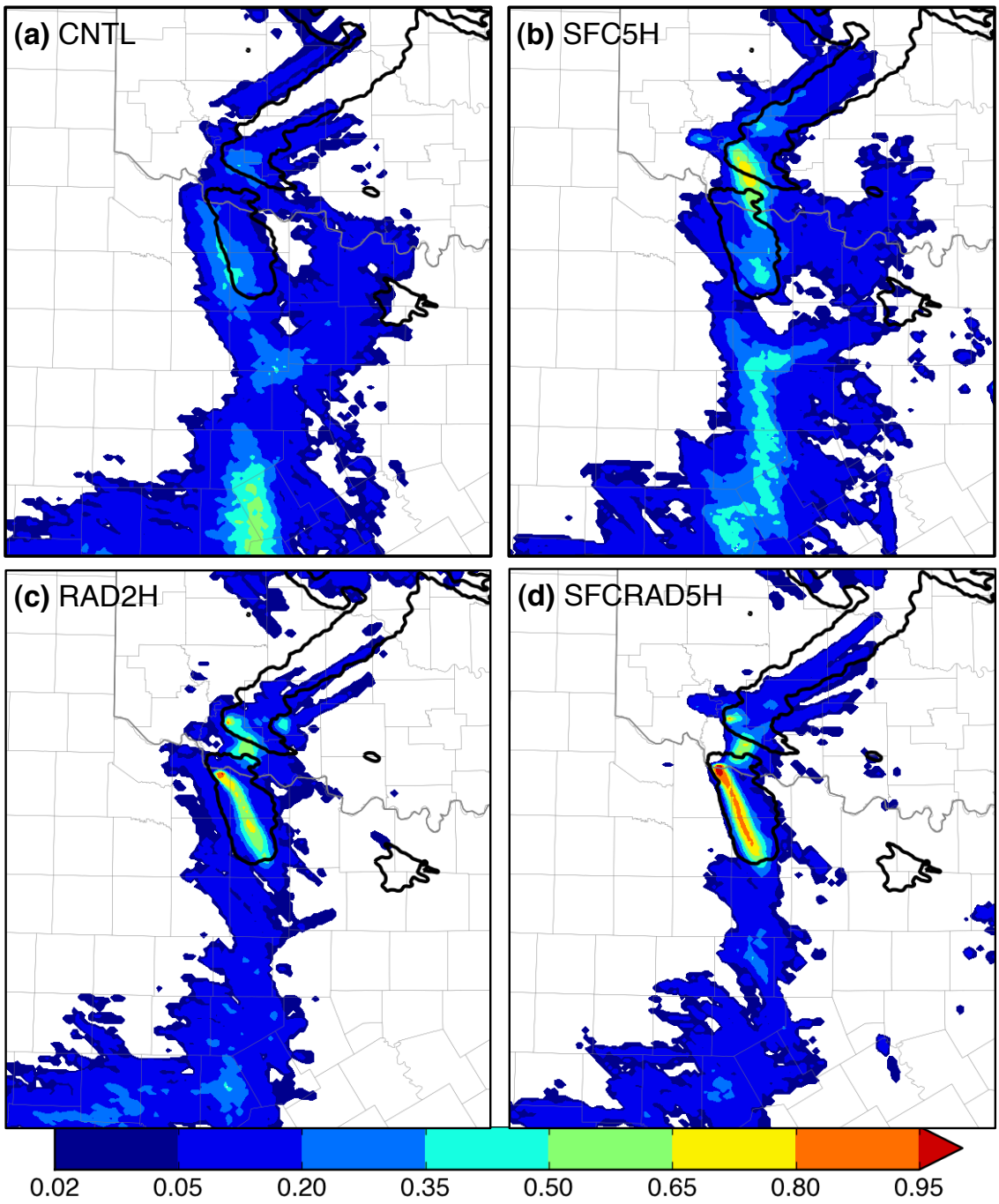

Figure 4.44: As in Fig. 4.42, but for PROB3H-UVV10 in the southern half of the domain.

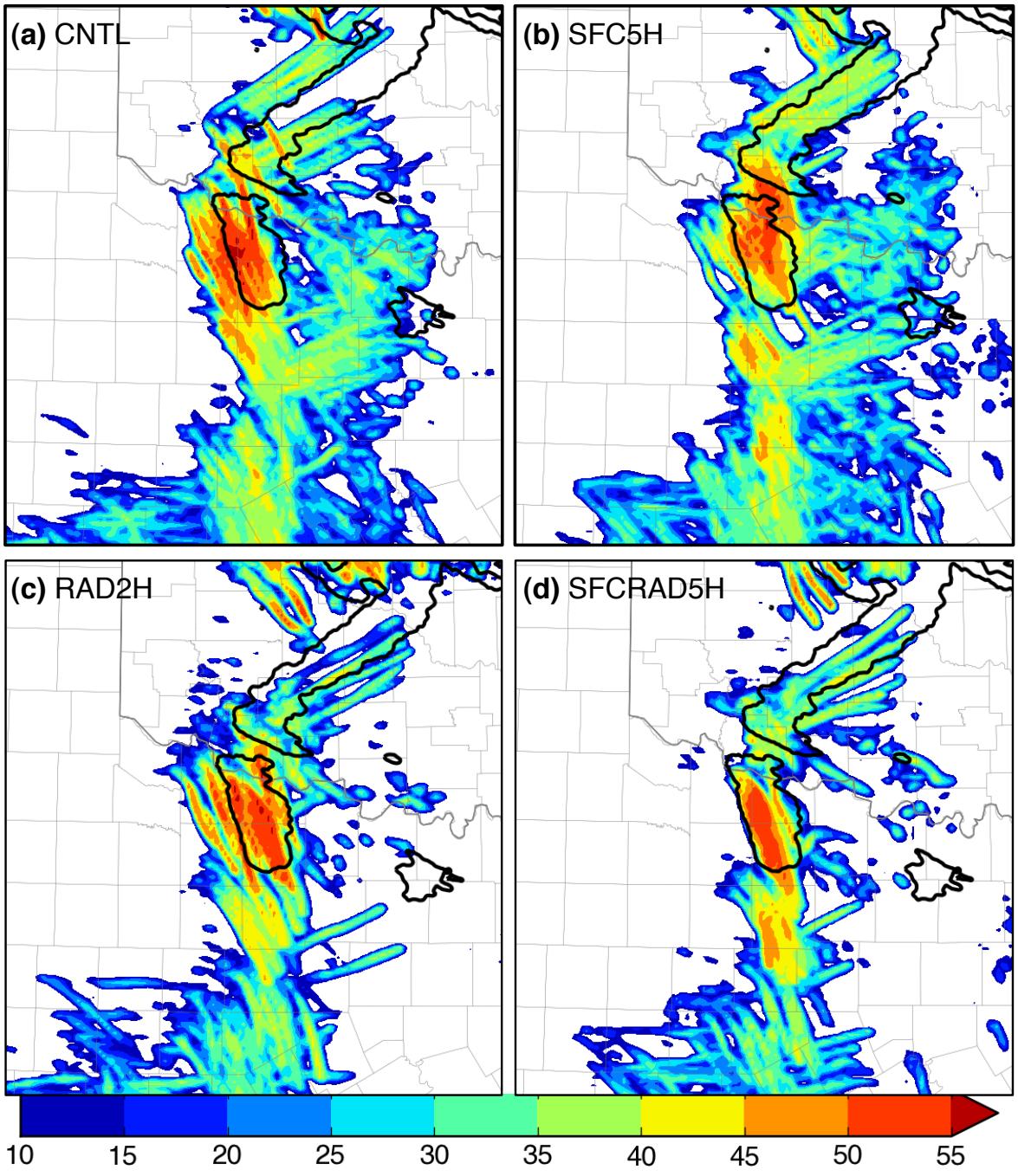

Figure 4.45: As in Fig. 4.43, but for MAX1H-UVV (m s^{-1}) in the southern half of the domain.

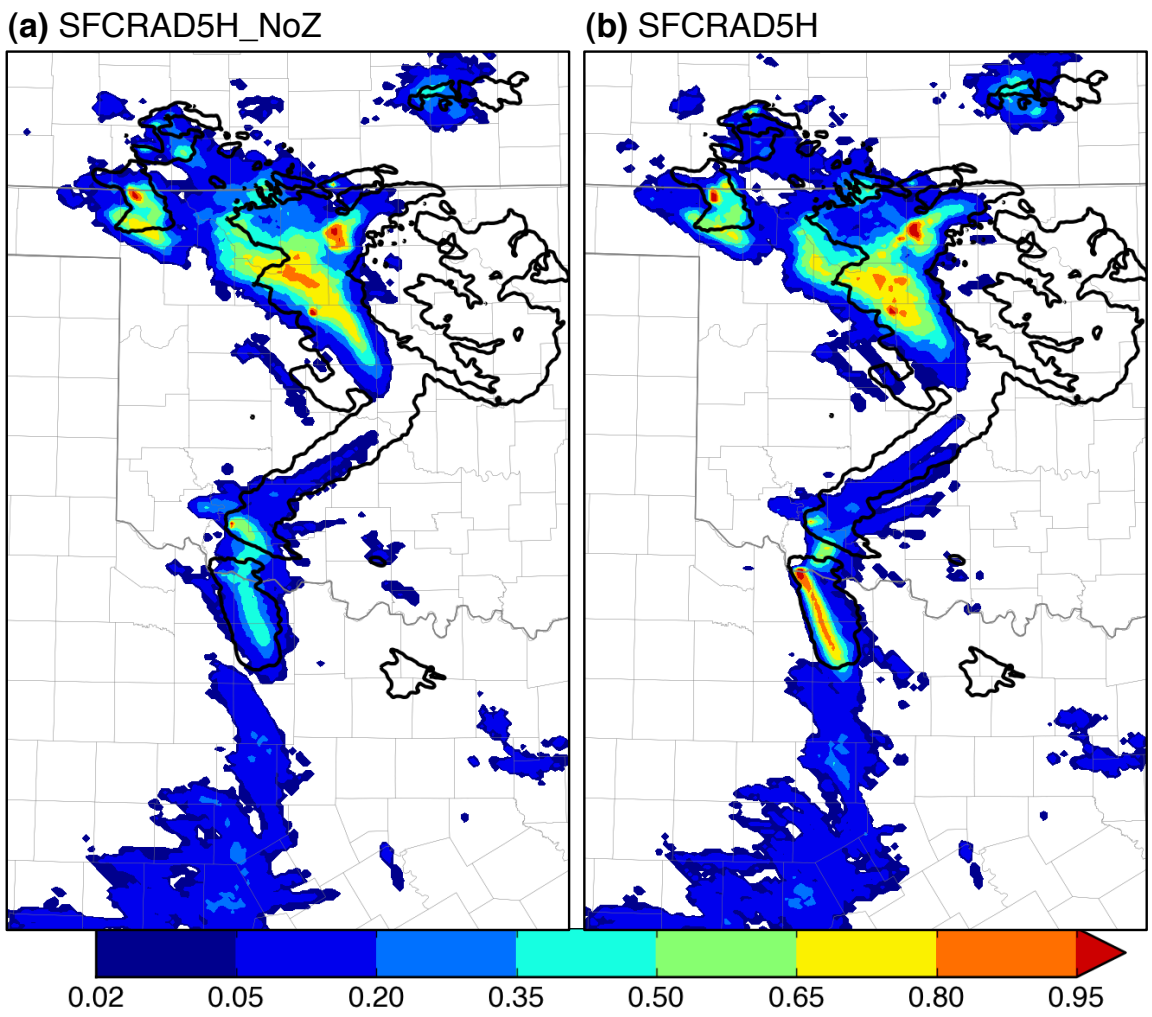

Figure 4.46: As in Fig. 4.38b, but for PROB3H-UVV10 from (a) SFCRAD5H-NoZ and (b) SFCRAD5H.

the posterior ensemble at 23 UTC (denoted SFCRAD5H-NoZ). Radial velocity and clear-air radar observations were assimilated and aside from the lack of reflectivity observations the experiment is designed identically to SFCRAD5H.

Reflectivity assimilation is especially beneficial for the forecast of the southern supercell originating within CI3 (c.f. Fig. 4.46a,b), while in other areas of the domain, differences between SFCRAD5H and SFCRAD5H-NoZ are less substantial. Without reflectivity assimilation, the maximum PROB3H-UVV10 values along the predicted track of the southern supercell in CI3 are reduced by 30-40%, and the probability

axis is shifted toward the east. Also, as indicated by low-end probabilities, several members produce convection to the west of the storm track, expanding the area of non-zero probabilities outside of the observed track. While reflectivity assimilation is beneficial for this storm, the forecast is slightly improved in other areas of the domain without reflectivity assimilation. For example, the PROB3H-UVV10 forecast in SFCRAD5H-NoZ produces a more narrow PROB3H-UVV10 axis in central OK that extends further to the southeast compared to SFCRAD5H, in slightly better correspondence with the observed track of convection in this corridor. Overall, the addition of reflectivity observations is not substantially detrimental to the PROB3H-UVV10 forecast, but does play a role in improving the forecast for the isolated convection in CI3.

During the second half of the DA period, the southern supercell was observed to weaken slightly and then rapidly strengthen near 23 UTC. The assimilation of reflectivity data may help spin-up the storm more quickly during this secondary intensification, improving SFCRAD5H over SFCRAD5H-NoZ. In addition, since much of the storm is between 5 – 15 km of the KFDR radar site, the portion of the storm above 5 km is only sampled by KTLX and KDYX, approximately 190 km and 200 km away from the storm, respectively. (at a range of 15 km the 19.5 degree beam, the highest radar tilt, is at 5 km AGL near the position of the southern supercell). At these ranges, the lowest tilts are spaced between 1.5 km and 2 km apart in the vertical, thus the impact of vertical correlations between radar data and the state is likely more important for this storm. Also, the large resolution volume for KTLX and KDYX radar observations likely leads to forward operator error at these ranges (the effects of the increased resolution volume are not accounted for in the forward operator). In other words, the reflectivity data from KFDR is important for the development of this storm by exploting the vertical correlations within the ensemble. This may help maintain the storm in the analyses as it passes near the

radar site, during a period of apparent weakening. This hypothesis will be further discussed in the following section.

4.6.7 Impact of radar localization on 0-3hr forecasts

In Part I, synthetic radar observations of a simulated convective system were assimilated into a collection of OSSEs. The OSSEs from Part I varied only in the values prescribed for the horizontal and vertical covariance localization used for the radar observations. The experiments with a horizontal localization cutoff of 18 km and vertical localization cutoff of 3 km both improved analysis quality among a set of experiments that systematically adjusted the cutoff length over a range of values. Yet, for all of the experiments conducted so far in this section, the localization cutoff was set to 12 km in the horizontal and 6 km in the vertical, more in-line with previous work since it was uncertain how the OSSE results would apply to a real-data case study. To assess the sensitivity of the forecasts to these localization choices, SFCRAD5H was repeated, but with the horizontal localization increased to 18 km (SFCRAD5H-H18V6). In another experiment, the vertical localization was decreased to 3 km (SFCRAD5H-H12V3).

Increasing the horizontal localization length produces several beneficial changes to the PROB3H-UVV10 forecast (c.f. Fig. 4.47b and Fig. 4.47c). Improvements include a larger area of probabilities >80% along the axis of forecast convection in central OK, and the elimination of the left-splits from CI3 in several members that were produced outside the observed track of this storm; the remaining members that produce a long-lived left-split do so along the observed path. Probabilities along the path of the southern supercell within CI3 are nearly unchanged between the two experiments. Decreasing the vertical localization length has a substantial negative impact on the forecast, primarily for the southern supercell in CI3 (c.f. Fig. 4.47a and Fig. 4.47b). Maximum PROB3H-UVV10 values associated

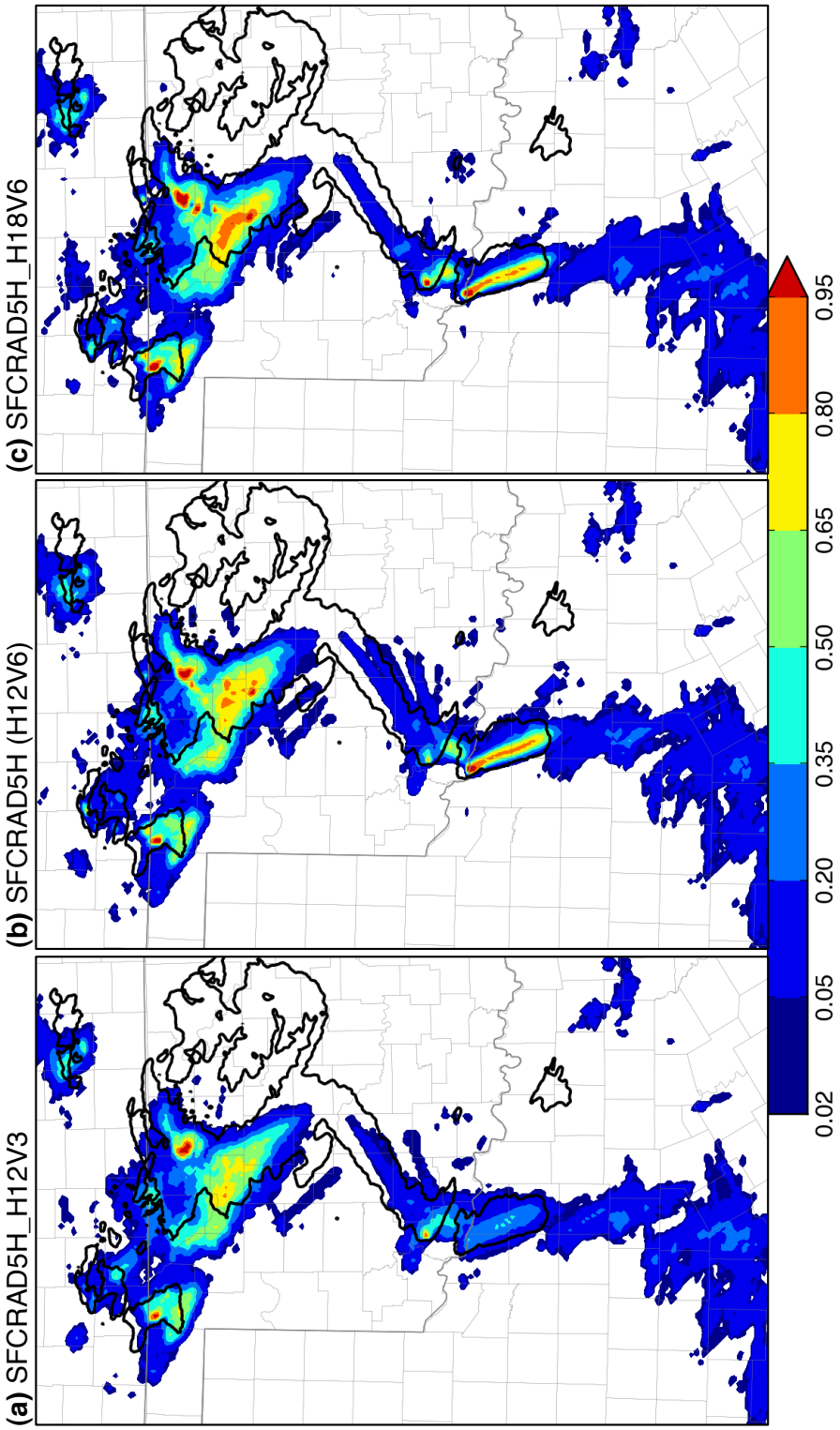

Figure 4.47: As in Fig. 4.38b, but for PROB3H-UVV10 from (a) SFCRAD5H-H12V3, (b) SFCRAD5H, and (c) SFCRAD5H-H18V6.

with this storm are decreased to <35%, and the probability axis is more diffuse (Fig. 4.47a). Forecast probabilities are also reduced in central OK, although the changes are smaller than those associated with the supercell in CI3.

These sensitivities suggest that permitting observations to update the state over a sufficiently deep layer can be important for producing accurate storm-scale analyses and short-term forecasts of convection in certain situations. In this case, the decrease of the vertical localization cutoff to 3 km in SFCRAD5H-H12V3 was detrimental to the analyses of convection for the supercell in CI3, as radar observations were unable to adequately update the state using the vertical covariance information. Given that the forecasts that did not assimilate reflectivity observations resulted in similar forecasts to those in this section that (section 4.6.6), the added benefit of deeper vertical localization is primarily provided by reflectivity observations. Further, more modest differences in other areas of the domain when the vertical cutoff is decreased or when reflectivity is not assimilated, suggest this is not a universal finding. As summarized in section 4.6.6, the reasons for this behavior may be due to both the precise pattern of sampling by the three nearest WSR-88Ds given the storm's position and the apparent weakening of convection at the end of the DA period.

4.6.8 Summary and Discussion

Results from several radar DA experiments were discussed in the present chapter. In the primary experiment, referred to as SFCRAD5H, radar and surface data were assimilated every 5-minutes during a two-hour period, beginning with the final set of SFC3H analyses at 21 UTC and ending at 23 UTC. Innovation diagnostics from SFCRAD5H revealed several biases in the prior ensemble analyses. These are summarized below, including a candidate explanation for the source of each error:

- Positive near-surface radial velocity bias due to model error during storm mergers.

- Positive near-surface reflectivity bias due to excessive size-sorting in the Morrison double-moment microphysics scheme.
- Positive mid-level reflectivity bias within the melting layer due to errors due to a simplified melting process in the forward operator.
- Negative upper-level reflectivity bias due to a slow bias in the environmental anvil-level wind or microphysical errors in the handling of ice in the anvil layer.

Little work has been undertaken to explicitly diagnose the systematic errors that lead to suboptimal updates during radar DA. As discussed in section 2.3.2.3, Dowell et al. (2011) identified various sources of error, many due to the inherent inability of the microphysics parameterization to represent certain physical processes, leading to detrimental effects on the analyses during DA. In their experiments with a single-moment microphysics scheme, large negative biases occurred below 3 km AGL, in contrast to the positive bias noted herein near the surface (c.f. Fig. 4.32b and their Fig. 4). The differences in near surface reflectivity bias between the present work and Dowell et al. (2011) are consistent with the deficiencies of single and double-moment microphysics schemes in producing reflectivity near the surface noted in other studies (e.g. Dawson et al. 2010; Kumjian and Ryzhkov 2012). A small positive bias is also evident in mid-levels in Fig. 4 of Dowell et al. (2011), yet this feature was not mentioned.

Using double-moment microphysics schemes and more complex forward operators during DA has only been conducted in a handful of studies. Jung et al. (2012), Putnam et al. (2013), and Yussouf et al. (2013) all noted improvements in the analyses due to the ability of double-moment schemes to better reproduce various features of supercells and mesoscale convective systems over single-moment schemes. Radar emulators, which have mostly been used to validate simulations of convective storms (e.g. Jung et al. (2010); Dawson et al. (2013), have the potential to be used

as a forward operator for DA (e.g. Jung et al. (2012)). Using the more advanced handling of melting ice in these emulators as part of the reflectivity forward operator would likely mitigate some of the bias present in the more simplified melting model used to compute reflectivity here.

Ensemble forecasts were initialized at 23 UTC from the final posterior SFCRAD5H analyses. The ensemble accurately predicted the short-term evolution of several regions of convection, including different types of convective modes. Various sensitivity experiments were conducted to assess the impact of surface and radar data (RAD2H and SFC5H), effectiveness of the reflectivity assimilation (SFCRAD5H-NoZ) and selection of horizontal and vertical localization for radar data (SFCRAD5H-H12V3 and SFCRAD5H-H18V6). A slow bias in the cloud-layer environmental wind profile in the analyses lead to timing errors in all of the forecasts. The axis of probabilities across central OK associated with the southeastward moving supercells, which produced the most intense severe weather, was largely insensitive to changes in the types of data assimilated (e.g. reflectivity) and variations in localization cutoff length. While part of the forecast probabilities in this region were due to convection that developed during the DA period within CI2, a larger amount of convection developed after 00 UTC. Since this occurs after the DA period, radar DA played a lesser role in 0-3hr forecast accuracy across the northern half of the domain. As discussed earlier, surface DA had a larger impact on the forecasts in this region, likely by adjusting surface boundaries (both mesoscale boundaries such as the dryline and convective-scale outflow boundaries associated with upscale growth), that led to more accurate short-term forecasts of convection.

In the southern part of the domain, the predicted convective evolution over the 3 hour forecast period in CI3 was more sensitive to the radar DA strategy. In experiments where radar data were withheld (SFC5H), low forecast probabilities were produced along the path of the observed southern supercell, while the northern storm

was forecast to persist as a long-lived right-moving supercell in a majority of ensemble members. When surface data were withheld (RAD2H), the forecast probability swath was aligned with the observed path, but some ensemble members develop convection to the west of the observed storm, reducing the magnitude of forecast probabilities. Assimilating both surface and radar DA provided the most accurate prediction of the track of the long-lived southern supercell during the three-hour forecast. The storm timing errors during the three-hour forecast seen in other areas were smaller for the long-lived southern supercell, likely due to the supercell dynamics causing a slower storm motion (Bunkers et al. 2000).

Forecasts of the left-split originating from the northern supercell in CI3 were less sensitive to radar or surface DA, although radar DA slightly reduced the probability values due to less robust convection associated with the northern supercell. Finally, the assimilation of reflectivity data and deeper vertical localization cutoff (6 km vs. 3 km) both played a role in producing a better initial representation of the right-moving supercell in CI3, leading to a substantially more accurate 0-3 hr forecast. The left-moving storm emanating from CI3 was less sensitive to these variations, although spurious left-splits were suppressed in the experiment with larger horizontal localization.

Chapter 5

Synthesis

Together, the OSSEs in chapter 3 and real-data experiments in chapter 4 provide two different means to assess various assimilation strategies and the impact of different datasets on convective-scale analyses and forecasts of developing convective systems. Several common themes emerged between the experiments within Part I and Part II. A synthesis of some of the key results of the dissertation are provided in this chapter.

Storm mergers and upscale growth In both the OSSEs and real-data experiments, periods of storm interactions and mergers were not analyzed or forecast well by the EnKF system. In the OSSEs, analysis errors increased during the initial period of mergers, while observation innovations increased during a storm merger within CI2 in the real-data experiments. As a result, improving forecasts of these types of transition events appears to pose a greater challenge compared to events composed of more isolated convection.

Radar data localization For both types experiments, increased horizontal localization for radar data improved the results, albiet modestly in the real-data experiments. Thus, the real-data experiments support the OSSE findings in Part I that a larger horizontal localization for radar data, up to 18 km, are not detrimental, and may be beneficial, to EnKF analyses of convection. Results were mixed to the sensitivity of vertical localization. In the OSSEs, a smaller vertical localization reduced analysis error, while this change was detrimental to the 29 May 2012 forecasts, primarily in the prediction of an isolated supercell. It is evident that there are situations when larger vertical localization is beneficial to analyses of convection,

such as when the storm is spinning up in the analyses and/or when a storm is not ideally placed with respect to observational assets (e.g. if a storm is very near to the radar, the upper parts of the storm may be more poorly observed). With regards to the vertical localization, the real-data experiments did not support the OSSE findings, and thus use of a vertical localization around 6 km remains an appropriate choice for radar data originating from WSR-88Ds.

Importance of multi-scale data assimilation The current real-data experiments underscore the importance of multi-scale DA systems in the production of accurate short-term ensemble forecasts of convection out to three-hours. This has been demonstrated in previous work for isolated convection (e.g. Stensrud and Gao 2010), and is reinforced the present work for a significantly more complex case. Both the forecasts of CI in section 4.5, and the forecasts of convective evolution in section 4.6, exhibited sensitivity to the mesoscale environment, with improved forecasts as the mesoscale environment was represented more realistically. Adjustments to the mesoscale environment through surface DA were able to improve convective forecasts by adjusting surface features and reducing surface moisture biases, while errors in the cloud-layer wind speed led to storm motions that were slower than observed and led for fairly large anvil-level reflectivity biases. For complex convective evolutions, these types of errors can have non-linear effects on the resulting forecast, but can be ameliorated by assimilating both meso- and convective-scale observations, thus embedding convection within an accurate mesoscale environment.

Using data assimilation to diagnose errors DA was used both to create forecast initial conditions and to isolate and estimate the magnitude of errors of various components of the DA system, including model error, forward operator error, instrument errors, and mesoscale environment error. For example, in the surface DA experiments, a positive moisture bias was present in the analyses due to a

previously known surface moisture bias. In the experiments that assimilated radar observations, various biases were identified due to model error from parameterization of microphysics, forward operator error, and errors in the mesoscale environment. Knowledge of these errors and their impact on the analyses should guide future efforts to improve the DA system, ideally leading to better forecasts.

In section 4.6.2, errors were identified primarily through the use of innovation statistics. Advanced methods to utilize the innovation diagnostics on-line, during the DA process, have been proposed and their applications to convective-scale EnKF are worthy of future work (Dee and Da Silva 1998). The adaptive inflation algorithm employed within DART provides another mechanism for identifying systematic errors in the analyses. This algorithm produces inflation fields for each component of the model state. The algorithm will increase prior-state inflation values where observations routinely fall outside of the prior ensemble, inflating the prior spread before DA. Spread deficiencies can occur due to inhomogeneities in the observing network, but also if analyses or observations are biased, producing inappropriate magnitudes of ensemble variance (Anderson 2009).

To illustrate this, the domain-maximum prior inflation values (posterior inflation is not employed in any of the present experiments) were computed in SFCRAD5H for 8 state fields (U10, V10, T2, Q2, QRAIN, QSNOW, QGRAUP, and REFL). At the surface, Q2 has a larger value of domain-maximum prior inflation than the other surface fields for most of the first three-hours of the DA period (Fig. 5.1). Afterward, the domain maximum Q2 prior inflation is comparable to the values for other surface fields, likely due to the development of convection after 21Z (since this is a domain maximum, the largest values of inflation are likely associated with convectively-generated cold pools). An increase in the prior inflation for all the surface fields occurs after 22 UTC associated with a large increase in convection within the northern half of the domain. Interestingly, during this period, the Q2

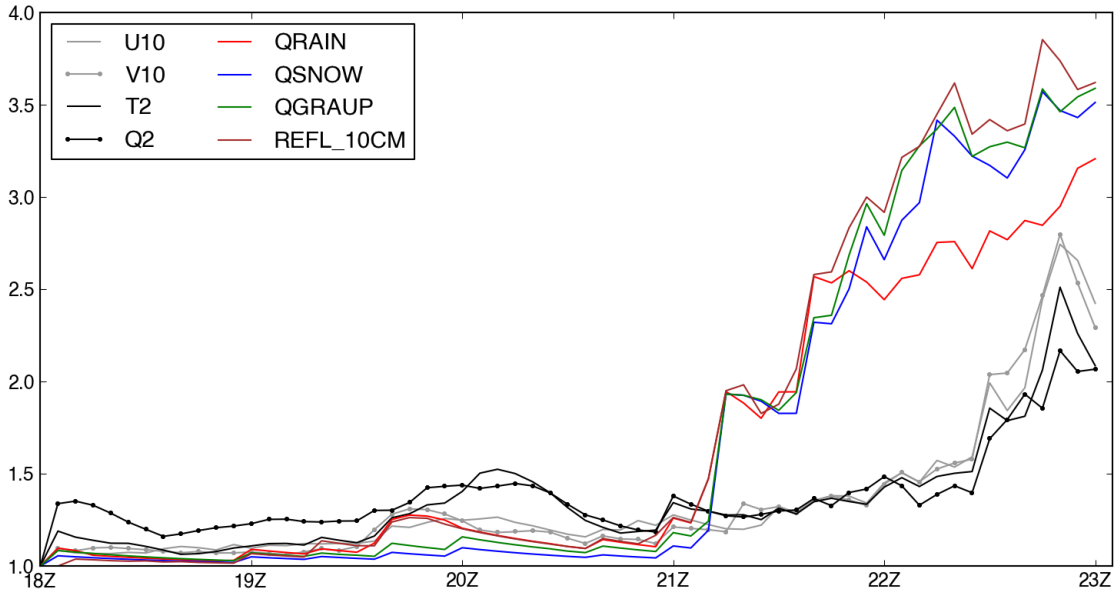

Figure 5.1: Domain-maximum inflation values for 8 state fields from SFCRAD5H.

domain-maximum prior inflation values are less than the other surface fields, with U10 and V10 having the largest values. This suggests errors in the low-level wind fields are larger than 2-m temperature and moisture errors within the developing cold pools.

The 4 microphysical-related inflation variables increase with the onset of convection shortly after 21 UTC. During the first 45 minutes, the domain-maximum values are similar, but after 21:45 UTC, the domain maximum prior inflation stabilizes for QRAIN, while the values for QSNOW and QGRAUP continue to increase. Greater uncertainty exists in the prediction of the ice species compared to rainwater (likely due to both model error and forward operator error), and adaptive inflation accounts for this by inflating QSNOW and QGRAUP more than QRAIN. In this way, the behavior of the inflation algorithm is able to provide estimates of relative error between state fields, providing additional information to complement the observation-space innovation diagnostics. Examination of the spatial distribution and other aggregated statistics for inflation will likely provide further understanding on the nature of model error. This task is left for future work.

Implications for warn-on-forecast systems All of the above factors have implications in proposed future warning systems that use short-term forecasts [$O(1hr)$] from ensembles of convective-allowing simulations to provide intensity, track, and uncertainty guidance for forecasters (Stensrud et al. 2009a, 2013). While the model and observation resolutions in the present set of experiments are too coarse to fully resolve the processes that forecasters are most interested in (e.g. low-level mesocyclones, downdrafts), the one-hour forecasts of convection in SFCRAD5H were generally quite good. Even beyond one-hour of lead-time, the forecasts exhibited skill at providing valuable track information for individual storms (e.g. SFCRAD5H provided an extremely accurate forecast of the southern supercell in CI3). Yet, even at higher resolutions, the forecasts of track and intensity will remain sensitive to the mesoscale environment in which convection is embedded. Further, model errors will play a role in the successful assimilation of surface and radar DA in any simulation that makes use of parameterized processes important to convection.

Bibliography

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824.

—, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Mon. Wea. Rev.*, **138**, 1273–1292.

Albers, S. C., J. A. McGinley, D. L. Birkenheuer, and J. R. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analyses of clouds, precipitation, and temperature. *Wea. Forecasting*, **11**, 273–287.

Ancell, B. C., 2012: Examination of analysis and forecast errors of high-resolution assimilation, bias removal, and digital filter initialization with an ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 3992–4004.

Ancell, B. C., C. F. Mass, and G. J. Hakim, 2011: Evaluation of surface analyses and forecasts with a multiscale ensemble Kalman filter in regions of complex terrain. *Mon. Wea. Rev.*, **139**, 2008–2024.

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, and R. Torn, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

—, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.

—, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83.

—, 2012: Localization and sampling error correction in ensemble kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371.

Anderson, J. L. and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463.

Armijo, L., 1969: A theory for the determination of wind and precipitation velocities with Doppler radars. *J. Atmos. Sci.*, **26**, 570–573.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor. Climatol.*, **3**, 396–409.

Bishop, C. H. and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus*, **61A**, 84–96.

—, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus*, **61A**, 97–111.

Blahak, U., 2007: RADAR_MIE_LM and RADAR_MIELIB – calculation of radar reflectivity from model output. Technical report, Internal Rep., Institute for Meteorology and Climate Research, University/Research Center Karlsruhe, 150 pp.

Bouttier, F., 1994: A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.*, **122**, 2376–2390.

Brandes, E. A., 1977: Flow in severe thunderstorms observed bu dual-Doppler radar. *Mon. Wea. Rev.*, **105**, 113–120.

—, 1984: Relationships between radar-derived thermodynamic variables and tornadogenesis. *Mon. Wea. Rev.*, **112**, 1033–1052.

Bratseth, A. M., 1986: Statistical interpolation by means of successive corrections. *Tellus*, **38A**, 439–447.

Brewster, K., 1996: Application of a Bratseth analysis scheme including Doppler radar data. *Preprints, 15th Conf. On Weather Analysis and Forecasting, Norfolk, VA, August 1996*, 92–95.

—, 2002: Recent advances in the diabatic initialization of a non-hydrostatic numerical model. *21st Conf. on Severe Local Storms*, San Antonio, TX, j51-54.

Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15**, 61–79.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094.

Chen, Y. and D. S. Oliver, 2009: Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput. Geosci.*, **14**, 579–601.

Clark, A. J., W. A. G. Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140.

—, 2010: Growth of spread in convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **25**, 594–612.

Coniglio, M. C., 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862.

Coniglio, M. C., D. J. Stensrud, and L. J. Wicker, 2006: Effects of upper-level shear on the structure and maintenance of strong quasi-linear mesoscale convective systems. *J. Atmos. Sci.*, **63**, 1231–1252.

Crook, A. and J. D. Tuttle, 1994: Numerical simulations initialized with radar-derived winds. Part II: Forecasts of three gust-front cases. *Mon. Wea. Rev.*, **122**, 1204–1217.

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767–1785.

Dawson, D. T., E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2013: Low-level zdr signatures in supercell forward flanks: the role of size sorting and melting of hail. *J. Atmos. Sci.*, in press.

Dawson, D. T., L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the environmental low-level wind profile on ensemble forecasts of the 4 May 2007 Greensburg, Kansas, tornadic storm and associated mesocyclones. *Mon. Wea. Rev.*, **140**, 696–716.

Dawson, D. T. and M. Xue, 2006: Numerical forecasts of the 15-16 June 2002 Southern Plains mesoscale convective system: Impact of mesoscale data and cloud analysis. *Mon. Wea. Rev.*, **134**, 1607–1629.

Dawson, D. T., M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.*, **138**, 1152–1171.

Dee, D. P. and A. M. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Met. Soc.*, **124**, 269–295.

Dong, J., M. Xue, and K. Droegemeier, 2011: The analysis and impact of simulated high-resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter. *Meteor. Atmos. Phys.*, **112**, 41–61.

Dowell, D. C. and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927.

Dowell, D. C., L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 oklahoma city supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 arcadia, oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophy. Res.*, **99 (C5)**, 10143–10162.

Fisher, R. A., 1915: Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507–521.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846–1868.

Gal-Chen, T., 1978: A method for the initialization of the anelastic equations: Implications for matching models with observations. *Mon. Wea. Rev.*, **106**, 587–606.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields*. Israel Program for Scientific Translation.

Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469.

Gao, J., M. Xue, A. Shapiro, and K. K. Droegemeier, 1999: A variational method for the analysis of three-dimensional wind fields from two Doppler radars. *Mon. Wea. Rev.*, **127**, 2128–2142.

Gao, J., M. Xue, A. Shapiro, Q. Xu, and K. K. Droegemeier, 2001: Three-dimensional simple adjoint velocity retrievals from single-Doppler radar. *J. Atmos. Oceanic Technol.*, **18**, 26–38.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Met. Soc.*, **125**, 723–757.

Ge, G., J. Gao, and M. Xue, 2012: Diagnostic pressure equation as a weak constraint in a storm-scale three-dimensional variational radar data assimilation system. *J. Atmos. Oceanic Technol.*, **29**, 1075–1092.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522.

Hacker, J. P. and D. Rostkier-Edelstein, 2007: PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon. Wea. Rev.*, **135**, 2958–2972.

Hacker, J. P. and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized pbl. *Mon. Wea. Rev.*, **133**, 3260–3275.

Hamill, T. M. and C. Snyder, 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552–1572.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Hane, C. E. and B. C. Scott, 1978: Temperature and pressure perturbations within convective clouds derived from detailed air motion information: Preliminary testing. *Mon. Wea. Rev.*, **106**, 645–661.

Hane, C. E., R. B. Wilhelmson, and T. Gal-Chen, 1981: Retrieval of thermodynamic variables within deep convective clouds: Experiments in three dimensions. *Mon. Wea. Rev.*, **109**, 564–576.

Hoffman, R. N. and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118.

Hong, S.-Y. and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

—, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hu, M. and M. Xue, 2007: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Wea. Rev.*, **135**, 507–525.

Hu, M., M. Xue, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorm. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675–698.

—, 2006b: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorm. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 675–698.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophy. Res.*, **113**, D13103.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press.

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 2930 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric radar data. *Mon. Wea. Rev.*, **140**, 1457–1475.

Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of a supercell storm using a two-moment bulk microphysics scheme. *J. Appl. Meteor. Climatol.*, **49**, 146–163.

Jung, Y., G. Zhang, and M. Xue, 2008: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, **136**, 2228–2245.

Kain, J. S., M. C. Coniglio, J. Correia, A. J. Clark, P. T. Marsh, C. L. Ziegler, V. Lakshmanan, S. D. M. Jr., S. R. Dembek, S. J. Weiss, F. Kong, M. Xue, R. A. Sobash, A. R. Dean, I. L. Jirak, and C. J. Melick, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.*, **94**, 1213–1225.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542.

Kain, J. S., I. L. Jirak, S. J. Weiss, A. J. Clark, M. C. Coniglio, J. Correia, Jr., A. R. Dean, P. T. Marsh, C. J. Melick, S. D. Miller, R. Sobash, M. Xue, F. Kong, K. W. Thomas, V. Lakshmanan, D. A. Imy, and S. R. Dembek, 2012: An overview of the 2012 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. *26th Conference on Severe Local Storms, Nashville, TN*.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic. Eng.*, **82D**, 35–45.

Kalman, R. E. and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic. Eng.*, **83**, 95–108.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.

Klemp, J. B. and R. B. Wilhemson, 1978: Simulations of right- and left-moving storms produced through storm splitting. *J. Atmos. Sci.*, **35**, 1097–1110.

Knopfmeier, K. H. and D. J. Stensrud, 2013: Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble Kalman filter. *Wea. Forecasting*, **28**, 815–841.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multiresoluton of ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparison of coarse- and fine-grid experiments. *Mon. Wea. Rev.*, **134**, 807–833.

Kumjian, M. R. and A. V. Ryzhkov, 2012: The impact of size sorting on the polarimetric radar variables. *J. Atmos. Sci.*, **69**, 2042–2060.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Lewis, J. M., S. Lakshmivarahan, and S. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press.

Lilly, D. K., 1990: Numerical prediction of thunderstorms - has its time come? *Quart. J. Roy. Met. Soc.*, **116**, 779–798.

Lin, X. and K. G. Hubbard, 2004: Uncertainties of derived dewpoint temperature and relative humidity. *J. Appl. Meteor. Climatol.*, **43**, 821–825.

Lin, Y., P. S. Ray, and K. W. Johnson, 1993: Initialization of a modeled convective storm using Doppler radar–derived fields. *Mon. Wea. Rev.*, **121**, 2757–2775.

Liu, H. and M. Xue, 2008: Prediction of convective initiation and storm evolution on 12 June 2002 during IHOP_2002. Part I: Control simulation and sensitivity experiments. *Mon. Wea. Rev.*, **136**, 2261–2282.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Met. Soc.*, **112**, 1177–1194.

Lorenc, A. C. and F. Rawlins, 2005: Why does 4D-Var beat 3D-Var? *Quart. J. Roy. Met. Soc.*, **131**, 3247–3257.

Lorenz, E., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333.

—, 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

Lu, H. and Q. Xu, 2009: Trade-offs between measurement accuracy and resolutions in configuring phased-array radar velocity scans for ensemble-based storm-scale data assimilation. *J. Appl. Meteor. Climatol.*, **48**, 1230–1244.

Markowski, P., Y. Richardson, E. Rasmussen, J. Straka, R. Davies-Jones, and R. J. Trapp, 2008: Vortex lines within low-level mesocyclones obtained from pseudo-dual-Doppler radar observations. *Mon. Wea. Rev.*, **136**, 3513–3535.

Marquis, J., Y. Richardson, P. Markowski, D. Dowell, J. Wurman, K. Kosiba, P. Robinson, and G. Romine, 2014: An investigation of the Goshen County, Wyoming, tornadic supercell of 5 June 2009 using EnKF assimilation of mobile mesonet and radar observations collected during VORTEX2. Part I: Experiment design and verification of the EnKF analyses. *Mon. Wea. Rev.*, in press.

Maxwell-Garnett, J. C., 1904: Colors in metal glasses and in metallic films. *Philos. Trans. Roy. Soc. London*, **A203**, 385–420.

Melhauser, C. and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371.

Meng, Z. and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423.

—, 2008a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540.

—, 2008b: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Wea. Rev.*, **136**, 3671–3682.

Michel, Y., T. Aulign, and T. Montmerle, 2011: Heterogeneous convective-scale background error covariances with the inclusion of hydrometeor variables. *Mon. Wea. Rev.*, **139**, 2994–3015.

Milbrandt, J. A. and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064.

Miller, P. A., M. F. Barth, and L. A. Benjamin, 2005: An update on MADIS observation ingest, integration, quality control, and distribution capabilities. *21st International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, San Diego, CA, j7.12.

Morrison, H., S. A. Tessendorf, K. Ikeda, and G. Thompson, 2012: Sensitivity of a simulated midlatitude squall line to parameterization of raindrop breakup. *Mon. Wea. Rev.*, **140**, 2437–2460.

Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007.

NCDC, 2012: Storm Data. Technical report, NOAA, 598 pp.

O'Brien, J. J., 1970: Alternative solutions to the classical vertical velocity problem. *J. Appl. Meteor. Climatol.*, **9**, 197–203.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1992: Ensemble prediction. *Proc. Seminar on Validation of Models over Europe*, 21–66.

Pereira, M. B. and L. Berre, 2006: The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev.*, **134**, 2466–2489.

Pu, Z., H. Zhang, and J. Anderson, 2013: Ensemble Kalman filter assimilation of near-surface observations over complex terrain: comparison with 3DVAR for short-range forecasts. *Tellus*, **65A**, 19620.

Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2013: The analysis and prediction of microphysical states and polarimetric radar variables in a mesoscale convective system using double-moment microphysics, multi-network radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, in press.

Ray, P. S., R. J. Doviak, G. B. Walker, D. Sirmans, J. Carter, and B. Bumgarner, 1975: Dual-Doppler observation of a tornadic storm. *J. Appl. Meteor. Climatol.*, **14**, 1521–1530.

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284.

Roux, F., 1985: Retrieval of thermodynamic fields from multiple-Doppler radar data using the equations of motion and the thermodynamic equation. *Mon. Wea. Rev.*, **113**, 2142–2157.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875–883.

Schenkman, A. D., M. Xue, and A. Shapiro, 2011a: Impact of CASA radar and Oklahoma Mesonet data assimilation on the analysis and prediction of tornadic mesovortices in an MCS. *Mon. Wea. Rev.*, **139**, 3422–3445.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011b: The analysis and prediction of the 8-9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139**, 224–246.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M. C. Coniglio, and M. S. Wandishin, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.

Shapiro, A., S. Ellis, and J. Shaw, 1995: Single-Doppler velocity retrievals with Phoenix II data: Clear air and microburst wind retrievals in the planetary boundary layer. *J. Atmos. Sci.*, **52**, 1265–1287.

Shapiro, A. and J. J. Mewes, 1999: New formulations of dual-Doppler wind analysis. *J. Atmos. Oceanic Technol.*, **16**, 782–792.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2008: A description of the Advanced Research WRF version 3. Technical report, NCAR Tech. Note NCAR/TN-475+STR, 113 pp.

Snook, N., M. Xue, and Y. Jung, 2011: Analysis of a tornadic mesoscale convective vortex based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, **139**, 3446–3468.

—, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146.

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.

Sobash, R. A. and D. J. Stensrud, 2013: The impact of covariance localization for radar data on EnKF analyses of a developing MCS: Observing system simulation experiments. *Mon. Wea. Rev.*, **141**, 3691–3709.

Stensrud, D. J. and J. Gao, 2010: Importance of horizontally inhomogeneous environmental initial conditions to ensemble storm-scale radar data assimilation and very short-range forecasts. *Mon. Wea. Rev.*, **138**, 1250–1272.

Stensrud, D. J., L. J. Wicker, K. E. Kelleher, M. Xue, M. P. Foster, J. T. Schaefer, R. S. Schneider, S. G. Benjamin, S. S. Weygandt, J. T. Ferree, and J. P. Tuell, 2009a: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Stensrud, D. J., L. J. Wicker, M. Xue, D. T. Dawson II, N. Yussouf, D. M. Wheatley, T. E. Thompson, N. A. Snook, T. M. Smith, A. D. Schenkman, C. K. Potvin, E. R. Mansell, T. Lei, K. M. Kuhlman, Y. Jung, T. A. Jones, J. Gao, M. C. Coniglio, H. E. Brooks, and K. A. Brewster, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16.

Stensrud, D. J., N. Yussouf, D. C. Dowell, and M. C. Coniglio, 2009b: Assimilating surface data into a mesoscale model ensemble: Cold pool analyses from spring 2007. *Atmos. Res*, **93**, 207–220.

Sun, J., 2005: Convective-scale assimilation of radar data: Progress and challenges. *Quart. J. Roy. Met. Soc.*, **131**, 3439–3463.

Sun, J. and A. Crook, 1994: Wind and thermodynamic retrieval from single-Doppler measurements of a gust front observed during Phoenix II. *Mon. Wea. Rev.*, **122**, 1075–1091.

Sun, J. and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.*, **54**, 1642–1661.

—, 1998: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part II: Retrieval experiments of an observed Florida convective storm. *J. Atmos. Sci.*, **55**, 835852.

Sun, J., D. W. Flicker, and D. K. Lilly, 1991: Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci.*, **48**, 876–890.

Sun, J., S. B. Trier, Q. Xiao, M. L. Weisman, H. Wang, Z. Ying, M. Xu, and Y. Zhang, 2012: Sensitivity of 0-12-h warm-season precipitation forecasts over the central United States to model initialization. *Wea. Forecasting*, **27**, 832–855.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I:theory. *Quart. J. Roy. Met. Soc.*, **113**, 1311–1328.

Tanamachi, R. L., L. J. Wicker, D. C. Dowell, H. B. Bluestein, D. T. I. Dawson, and M. Xue, 2013: EnKF assimilation of high-resolution, mobile Doppler radar data of the 4 May 2007 Greensburg, Kansas, supercell into a numerical cloud model. *Mon. Wea. Rev.*, **141**, 625–648.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the rapid update cycle. *Wea. Forecasting*, **18**, 1243–1261.

Thompson, T. E., L. J. Wicker, and X. Wang, 2012: Impact from a volumetric radar-sampling operator for radial velocity observations within EnKF supercell assimilation. *J. Atmos. Oceanic Technol.*, **29**, 1417–1427.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.

Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generations of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Tracton, M. S. and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 379–398.

Tremolet, Y., 2004: Diagnostics of linear and incremental approximations in the 4D-Var. *Quart. J. Roy. Met. Soc.*, **130**, 2233–2251.

Wakimoto, R. M., 1982: The life cycle of thunderstorm gust fronts as viewed with Doppler radar and rawinsonde data. *Mon. Wea. Rev.*, **110**, 1060–1082.

Wang, H., J. Sun, X. Zhang, X.-Y. Huang, and T. Augligne, 2013: Radar data assimilation with WRF 4D-Var. Part I: System development and preliminary testing. *Mon. Wea. Rev.*, **141**, 2224–2244.

Wang, Q.-W. and M. Xue, 2012: Convective initiation on 19 June 2002 during IHOP: High-resolution simulations and analysis of the mesoscale structures an- convection initiation. *J. Geophy. Res.*, **117**, D12107.

Weckwerth, T. M., 2000: The effect of small-scale moisture variability on thunderstorm initiation. *Mon. Wea. Rev.*, **128**, 4017–4030.

Weckwerth, T. M. and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.*, **134**, 5–22.

Weckwerth, T. M., D. B. Parsons, S. E. Koch, J. A. Moore, M. A. LeMone, B. B. Demoz, C. Flamant, B. Geerts, J. Wang, and W. F. Feltz, 2004: An overview of the International H2O project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.*, **85**, 253–277.

Weisman, M. L. and R. Rotunno, 2004: "A theory for strong long-lived squall lines" revisited. *J. Atmos. Sci.*, **61**, 361–382.

Weygandt, S. S., A. Shapiro, and K. K. Droegemeier, 2002: Retrieval of model initial fields from single-Doppler observations of a supercell thunderstorm. Part I: Single-Doppler velocity retrieval. *Mon. Wea. Rev.*, **130**, 433–453.

Wheatley, D. M. and D. J. Stensrud, 2010: The impact of assimilating surface pressure observations on severe weather events in a WRF mesoscale ensemble system. *Mon. Wea. Rev.*, **138**, 1673–1694.

Wheatley, D. M., D. J. Stensrud, D. C. Dowell, and N. Yussouf, 2012: Application of a WRF mesoscale data assimilation system to springtime severe weather events 200709. *Mon. Wea. Rev.*, **140**, 1539–1557.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Wu, B., J. Verlinde, and J. Sun, 2000: Dynamical and microphysical retrievals from Doppler radar observations of a deep convective cloud. *J. Atmos. Sci.*, **57**, 262–283.

Wurman, J., Y. Richardson, C. Alexander, S. Weygandt, and P. F. Zhang, 2007: Dual-Doppler and single-Doppler analysis of a tornadic storm undergoing mergers and repeated tornadogenesis. *Mon. Wea. Rev.*, **135**, 736–758.

Xiao, Q. and J. Sun, 2007: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP_2002. *Mon. Wea. Rev.*, **135**, 3381–3404.

Xu, Q., H. Lu, S. Gao, M. Xue, and M. Tong, 2008: Time-expanded sampling for ensemble Kalman filter: Assimilation experiments with simulated radar observations. *Mon. Wea. Rev.*, **136**, 2651–2667.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161–193.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, 1995: ARPS Version 4.0 User's Guide. [Available at http://www.caps.ou.edu/ARPS], 380 pp.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D. Wang, 2001: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143–165.

Xue, M., M. Hu, and A. Schenkman, 2013: Numerical prediction of 8 May 2003 Oklahoma City tornadic supercell and embedded tornado using ARPS with the assimilation of WSR-88D radar data. *Wea. Forecasting*, in press.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *24th Conf. Several Local Storms*, Savannah, GA, page 12.2.

Xue, M. and W. J. Martin, 2006a: A high-resolution modeling study of the 24 May 2002 dryline case during IHOP. Part I: Numerical simulation and general evolution of the dryline and convection. *Mon. Wea. Rev.*, **134**, 149–171.

—, 2006b: A high-resolution modeling study of the 24 May 2002 dryline case during IHOP. Part II: Horizontal convective rolls and convective initiation. *Mon. Wea. Rev.*, **134**, 172–191.

Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. *J. Atmos. Oceanic Technol.*, **23**, 46–66.

Xue, M., D. Wang, J. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteorol. Atmos. Phys.*, **82**, 139–170.

Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 oklahoma city tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412.

Zhang, F., Z. Meng, and A. Aksoy, 2006a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Wea. Rev.*, **134**, 722–736.

Zhang, F., A. M. Odins, and J. W. Neilsen-Gammon, 2006b: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125.

Zhang, J., F. Carr, and K. Brewster, 1998: ADAS cloud analysis. *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, 185–188.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. V. Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D.-J. Seo, E. Wells, and C. Dempsey, 2011: National mosaic and multi-sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338.

Zhao, Q., J. Cook, Q. Xu, and P. R. Harasti, 2008: Improving short-term storm predictions by assimilating both radar radial-wind and reflectivity observations. *Wea. Forecasting*, **23**, 373–391.

Ziegler, C. L., 1985: Retrieval of thermal and microphysical variables in observed convective storms. Part I: Model development and preliminary testing. *J. Atmos. Sci.*, **42**, 1487–1509.

—, 1988: Retrieval of thermal and microphysical variables in observed convective storms. Part II: Sensitivity of cloud processes to variation of the microphysical parameterization. *J. Atmos. Sci.*, **45**, 1072–1090.

Appendix A

2012 NCAR cycled mesoscale EnKF system

The NCAR mesoscale EnKF system was run in real-time during the spring of 2012. Observations were assimilated every 6 hours between 00 UTC 30 April 2012 and 02 July 2012, with a brief interruption due to a computing shutdown on 26-27 June. This shutdown is after the period of interest in this work, thus the assimilation system had been cycling continuously for approximately one month prior to the current case. Assimilated observations included rawinsondes, METAR, NOAA profilers, buoy, ship, AMDAR reports, atmospheric motion vectors, and GPS radio occultation observations (Fig. A.1). The domain covers most of the contiguous United States, the eastern Pacific Ocean, southern Canada, and the Gulf of Mexico (outer domain in Fig. 4.7). Lateral boundary conditions were provided by the operational GFS analysis and 6-hour forecast.

The following physics choices were used in the cycled system: Tiedtke cumulus parameterization, RRTMG radiation, and Morrison microphysics. Horizontal and vertical localization was set to 640 km and 8 km, respectively, using the Gaspari-Cohn localization function. In areas of dense observations, exceeding 2000 within the localization ellipsoid, the localization distance was decreased by the observation number overage ratio (if 4000 observations exist within the ellipsoid, the localization length was halved). A sampling error correction (Anderson 2012) was also applied during assimilation to reduce to effects from spurious covariances.

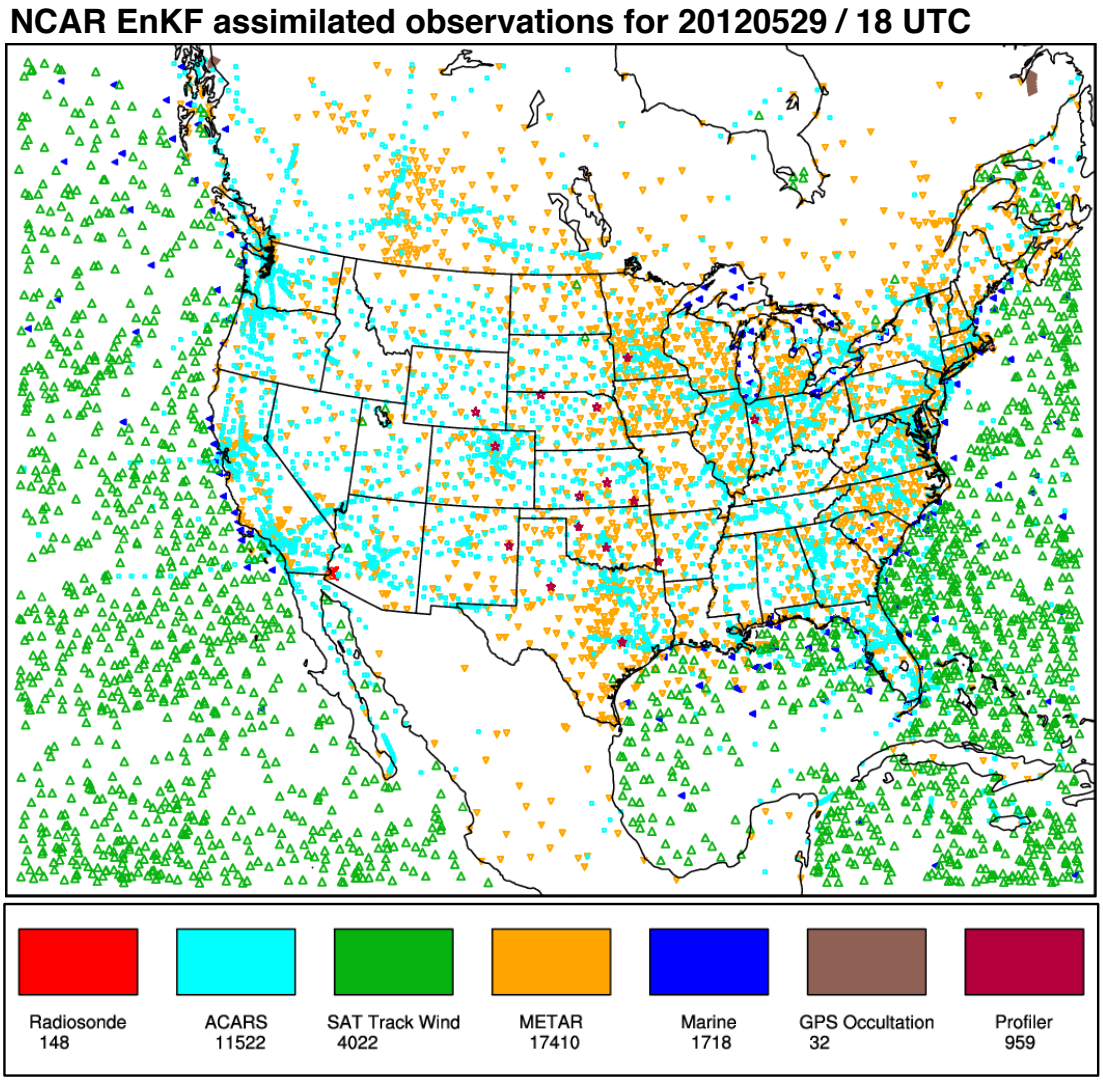

Figure A.1: Observations assimilated during the 18 UTC 2012 May 29 cycle of the NCAR EnKF mesoscale analysis system. Observations types and the number of assimilated observations are provided on the figure legend.

Appendix B

Forward operators for Vr, Z and Vt

Reflectivity is computed within the Morrison microphysics scheme and interpolated directly to produce the predicted observation values, $H(x^f)$, from the prior state. Reflectivity-weighted terminal velocity ($V_{T_x}^Z$) is also computed within the scheme and used by the forward operator for radial velocity. The computation for $V_{T_x}^Z$ was added to the Morrison scheme. Since a double-moment microphysics scheme is used, the form of the reflectivity forward operator is different than that of single-moment schemes (since number concentration is now predicted and not diagnosed). These formulations are summarized below.

Reflectivity for rain, snow, and graupel The rain, snow, and graupel particle size distributions (PSDs) are modelled as inverse exponential functions of the form:

$$N_x(D) = N_{0x} e^{-\lambda_x D} \tag{B.1}$$

Where x is a given hydrometeor species (rain, snow, or graupel), D is the hydrometeor diameter, $N_x(D)$ is the number concentration at a given particle diameter, N_{0x} is the number concentration intercept parameter, and λ_x is the slope parameter. Under the assumption of Rayleigh scattering, and that mass and diameter can be related in a power-law relationship of the form $m_x(D_x) = c_x D_x^{d_x}$, where c_x and d_x are constants, the radar reflectivity factor can be computed by integrating across the PSD for each species[1]:

$$Z_x = 720 N_{Tx} D_{nx}^6 \tag{B.2}$$

[1] An approachable derivation of this equation can be found in Milbrandt and Yau (2005)

where N_{T_x} is the total number concentration and D_{nx} is the characteristic diameter, or λ_x^{-1}, and is computed as:

$$D_{nx} = (\frac{\rho q_x}{c_x N_{T_x} \Gamma(1+d_x)})^{(\frac{1}{d_x})} \tag{B.3}$$

where q_x is the mixing ratio, ρ is the air density, and Γ is the gamma function. N_{T_x} and q_x are both predicted by the microphysics scheme. For spherical particles, $c_x = \frac{\pi}{6}\rho_x$ and $d_x = 3$, where ρ_x is the particle density. B.2 is used for rain, but is modified for graupel and snow to include the effects of ice on the backscattering cross-section through the following adjustment, assuming spherical particles,

$$Z_{eq} = \frac{|K|_i^2}{|K|_w^2}(\frac{\rho_x}{\rho_r})^2 Z_x \tag{B.4}$$

where $|K|_i$ is the dielectric constant for ice, $|K|_w$ is the dielectric constant for liquid water, ρ_r is the density of liquid water and ρ_x is the density of either snow or graupel. Z_{eq} is considered an *equivalent* radar reflectivity factor. The ratio of the dielectric constants $\frac{|K|_i^2}{|K|_w^2} = 0.23$.

Reflectivity for melting graupel and snow Large uncertainties exist in modeling the effects of melting graupel and snow on reflectivity. In most data assimilation studies (e.g. Dowell et al. 2011), the reflectivity of snow or graupel in an environment with temperature $>0^\circ$ C was formulated either similarly to rain, or under the assumption that the particle surface remained dry during the melting process (due to absorption into the particle interior).

The effects of melting graupel and snow on reflectivity in the Morrison scheme are based off the code of Blahak (2007). This code computes the dielectric constant for a melting particle using the Maxwell-Garnett mixing formula (Maxwell-Garnett 1904). To determine the degree of melting, the forward operator computes a meltwater fraction (MWF) as the particle descends through the melting layer. The MWF is defined as the fraction of mixing ratio that is present compared to the mixing

ratio at the level above the melting layer. This increases from 0 to 1 as a melting particle descends. Further, the forward operator assumes that 90% of the meltwater will remain on the surface while 10% is absorbed into the interior of the particle. Given these assumptions, the PSD is broken into bins, and the reflectivity for each bin due to melting snow or graupel is determined by computing the backscattering cross-section for the modeled ice-air-water mixture within the bin, using the Rayleigh assumption. These are summed over the PSD to determine the reflectivity for melting snow or melting graupel. This technique allows for a continuously varying particle density (as ice changes to rain, gradually filling air pockets within the ice lattice) and varying dielectric constant throughout the melting process.

Finally, the total equivalent reflectivity is computed as the sum of the reflectivity values from each species, followed by the conversion to a logarithmic scale. At grid points where the temperature is $<0^\circ$ C, equation (B.2) is used to compute reflectivity for rain, snow and graupel. At grid points where the temperature is $>0^\circ$ C, reflectivity for snow and graupel, if any exists, is computed following the procedure described in the previous paragraph.

$$Z_e = Z_{rain} + Z_{snow} + Z_{graupel} \tag{B.5}$$

$$Z_{dB} = 10log_{10}Z_e \tag{B.6}$$

Reflectivity-weighted terminal velocity Assuming the distribution in B.1, and that terminal velocity and diameter can be related in a power-law relationship of the form $V_{T_x}(D_x) = a_x D_x^{b_x}$, where a_x and b_x are constants, $V_{T_x}^Z$ is computed for rain, snow, and graupel using:

$$V_{T_x}^Z = a_x D_{nx}^{b_x} \frac{\Gamma(2d_x + b_x + 1)}{\Gamma(2d_x + 1)} (\frac{\rho_{850}}{\rho})^{0.54} \tag{B.7}$$

A density correction is applied to $V_{T_x}^Z$ via the air density ratio raised to the 0.54 power in the above equation. Further, $V_{T_x}^Z$ is not permitted to increase above 9.1 m s^{-1}, 1.2 m s^{-1}, and 20 m s^{-1} for rain, snow, and graupel, respectively.

A combined V_{T_Z} is computed as,

$$V_T^Z = \frac{V_{T_r}^Z Z_r + V_{T_s}^Z Z_s + V_{T_g}^Z Z_g}{Z_r + Z_s + Z_g} \tag{B.8}$$

Radial velocity The radial velocity is computed using interpolated values of the U, V, W, and V_T^Z fields, with knowledge of the radar beam elevation angle α and the azimuth angle β:

$$V_r = Ucos(\alpha)sin(\beta) + Vcos(\alpha)sin(\beta) + (W - V_T^Z)sin(\alpha) \tag{B.9}$$